NCT Number: NCT02912468

STATISTICAL ANALYSIS PLAN

A randomized, 24-week treatment, double-blind, placebo-controlled efficacy and safety study of dupilumab 300 mg every other week, in patients with bilateral nasal polyposis on a background therapy with intranasal corticosteroids

SAR231893(REGN668)-EFC14146

STATISTICIAN: [redacted]

DATE OF ISSUE: 14-Aug-2017

Total number of pages: 95

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

ACQ-6	Asthma Control Questionnaire 6-question version
AD	Atopic Dermatitis
ADA	Anti-drug Antibodies
AE	Adverse Event
AESI	Adverse Event of Special Interest
ALT	Alanine Aminotransferase
AM	Ante Meridiem
ANA	Antinuclear Antibodies
BID	Twice Daily
CI	Confidence Interval
CPK	Creatine Phosphokinase
CRFs	Case Report Forms
CT	Computed Tomography
CV%	Coefficient of Variation
DMC	Data Monitoring Committee
DNA	Deoxyribonucleic Acid
ECG	Electrocardiography
ECP	Eosinophil Cationic Protein
e-CRF	Electronic Case Report Form
e-diary	Electronic Diary
EQ-5D	European Quality of Life-5D Scale
EOS	End-of-Study
EOT	End-of-Treatment
FEF 25-75	Forced Expiratory Flow at 25% to 75% of forced vital capacity
FEV1	Forced Expiratory Volume in one second
FVC	Forced Vital Capacity
HBc Ab	Hepatitis B Core Antibody
HBs Ab	Hepatitis B Surface antibody
HBs Ag	Hepatitis B Surface Antigen
HBV	Hepatitis B Virus
HCV Ab	Hepatitis C Virus Antibody
HLGT	High-Level Group Term
HLT	High Level Term
HIV	Human Immunodeficiency Virus
HRQoL	Health related quality of life
ICH	International Council for Harmonisation
IDMC	Independent Data Monitoring Committee
IEC	Institutional Ethics Committee
IgE	Immunoglobulin E
IMP	Investigational Medicinal Product
INCS	Intranasal Corticosteroid Spray

INN	International Nonproprietary Name
IRB	Institutional Review Board
ITT	Intent-to-treat
IV	Intravenous
IRT	Interactive Response Technologies
IVRS/IWRS	Interactive Voice Response System/Interactive Web Response System
K-M	Kaplan-Meier
LFT	Liver Function Test
LLOQ	Lower Limit of Quantitation
LMK	Lund Mackay
LS	Least Squares
MCID	Minimal Clinically Important Difference
MFNS	Mometasone Furoate Spray
MI	Multiple Imputation
MID	Minimal Important Difference
MMRM	Mixed-effect Model with Repeated Measures
MRI	Magnetic Resonance Imaging
NC	Nasal Congestion
NERD	Nonsteroidal Anti-inflammatory Drug Exacerbated Respiratory Disease
NIMP	Noninvestigational Medicinal Product
NP	Nasal Polyposis
NPIF	Nasal Peak Inspiratory Flow
NPS	Nasal Polyp Score
NSAID	Nonsteroidal Anti-inflammatory Drug
OC	Osteomeatal Complex
PCSA	Potentially Clinically Significant Abnormalities
PD	Pharmacodynamics
PGDM	Metabolite of Prostaglandin D2
PK	Pharmacokinetics
PROs	Patient-Reported Outcomes
PT	Preferred Term
q2w	Every 2 weeks
q4w	Every 4 weeks
QD	Once daily administration
QoL	Quality of Life
qw	Once weekly
RNA	Ribonucleic Acid
SAE	Serious Adverse Event
SC	Subcutaneously
SCS	Systemic Corticosteroid
SD	Standard Deviation
SEM	Standard Error of the Mean
SNOT-22	Sinonasal Outcomes Test 22-item scale
SOC	System Organ Class
TARC	Thymus and Activation-Regulated Chemokine

TE	Treatment-Emergent
TEAE	Treatment-Emergent Adverse Event
Th2	Type 2 T-helper cell
TSS	Total Symptoms Score
ULN	Upper Limit of Normal
UPSIT	University of Pennsylvania Smell Identification Test
US	United States
V	Visit
VAS	Visual Analogue Scale
WBC	White Blood Cell
WOCF	Worst Observation Carried Forward

1 OVERVIEW AND INVESTIGATIONAL PLAN

1.1 STUDY DESIGN AND RANDOMIZATION

EFC14146 is phase 3 multinational, multicenter, randomized, double-blind, placebo-controlled, parallel group study that will evaluate the efficacy of dupilumab 300 mg administered subcutaneously (SC) q2w for 24 weeks followed by 24 weeks off-treatment period on a background of INCS. Both the patient and investigator will be blinded to the assigned drug, active or matching placebo in identical pre-filled syringes.

After a run-in period of 4 weeks ±3 days, patients will be centrally randomized using a permuted block randomization schedule via Interactive Voice Response System/Interactive Web Response System (IVRS/IWRS) in a 1:1 randomization ratio for dupilumab 300 mg q2w until Week 24 (Arm A) and placebo matching dupilumab q2w administration (Arm B). Randomization will be stratified by asthma/ non-steroid anti-inflammatory drug exacerbated respiratory disease (NERD) status (with co-morbid asthma and/or NERD history, without co-morbid asthma nor NERD history) at screening (V1), prior surgery (with prior surgery for NP, without any prior surgery for NP) at screening (V1) and country.

A total of approximately 240 patients with bilateral NP will be randomized to two treatment arms (120 patients/arm). Recruitment for patients without asthma/ NERD will be limited to 120 patients (out of the total of 240 randomized patients). Recruitment for patients without any prior surgery for nasal polyposis will be limited to 120 patients (out of the total 240 randomized patients). Patients may fall in more than one category without limitation in numbers.

1.2 OBJECTIVES

1.2.1 Primary objectives

The primary objective of this study is to evaluate the efficacy of dupilumab 300 mg every 2 weeks (q2w) compared to placebo on a background of Mometasone Furoate Nasal Spray (MFNS) in reducing nasal congestion/obstruction (NC) severity and endoscopic nasal polyp score (NPS) in patients with bilateral nasal polyposis (NP) at Week 24. In addition for regulatory purpose in Japan only, reduction in computed tomography (CT) scan opacification of the sinuses will be also a co-primary objective.

1.2.2 Secondary objectives

- To evaluate the efficacy of dupilumab in reducing CT scan opacification of the sinuses as assessed by the Lund Mackay (LMK) score (for regulatory purposes ex-Japan; in Japan, this is part of the primary objective).
- To evaluate the efficacy of dupilumab in improving total symptoms score (TSS).

- To evaluate the efficacy of dupilumab in improving sense of smell through daily loss of smell assessment and UPSIT functional smell test.
- To evaluate the effect of dupilumab on patient reported outcome (PRO) sino-nasal outcome test (SNOT-22).
- To evaluate the efficacy of dupilumab in reducing the proportion of patients requiring rescue treatment defined as: use systemic corticosteroids (SCS) or NP surgery (actual or planned) during the treatment period.
- To evaluate the efficacy of dupilumab on lung function (forced expiratory volume in one second, FEV1) in the sub-group of patients with asthma at Week 24.
- To evaluate the residual effect of dupilumab through the 24-week follow-up period.
- To evaluate the safety of dupilumab in patients with bilateral NP.
- To evaluate functional dupilumab concentrations (systemic exposure) and incidence of treatment-emergent anti-drug antibodies (ADA).

1.2.3 Exploratory objectives

- To explore the effects of dupilumab on nasal microbiome, total and allergen-specific IgEs in blood, and urinary leukotriene E4 and a metabolite of prostaglandin D2 (PGDM).
- To evaluate the effect of dupilumab on healthcare resource utilization.
- To evaluate the effect of dupilumab on SNOT-22 items: “decreased sense of smell/taste”, “difficulty falling asleep”, “wake up at night”, “lack of a good night's sleep”, “wake up tired”, “fatigue”, and “reduced productivity”.
- To evaluate the effect of dupilumab on health related quality of life (HRQoL) scale (Index Score of EQ5D-5L) and EQ-5D-5L VAS score (self-rated health).
- To assess the effect of dupilumab in improving sense of taste.

1.3 DETERMINATION OF SAMPLE SIZE

The sample size was chosen to enable an adequate characterization of the efficacy between dupilumab 300 mg q2w and placebo with regard to the 2 co-primary endpoints, changes from baseline in NC and NPS at Week 24.

- The observed mean difference in NC reduction between the dupilumab group with 300 mg weekly dosing (qw) and placebo in the proof of concept study of dupilumab in patients with bilateral nasal polyposis (ACT12340) was 0.95-0.26=0.69, and discounting 20% for 300 q2w, the mean difference was assumed to be 0.69*0.8=0.55. (NC scores range from 0 to 3)
- The observed common standard deviation (SD) was 0.86, and inflated by 20%, the assumed SD was 0.86*1.2=1.03.
- Then the assumed effect size was 0.55/1.03=0.534.

Assuming normal distribution of the change in NC, a common standard deviation (SD) of 1.03, and a 25% dropout rate, with 120 patients per group, the study will have 95% power to detect an effect size of 0.534 using a two-sided test with alpha = 0.05 for the change in NC at Week 24 in the dupilumab 300 mg q2w group.

- The observed mean NPS reduction of the dupilumab group with 300 mg qw dosing in the proof of concept study (ACT12340) was 1.85 and the observed mean NPS reduction of the placebo group was 0.30. So the observed mean difference is 1.85-0.30=1.55. Discounting 20% for 300 q2w, the mean difference was assumed to be 1.55*0.8= 1.24 (NPS values range from 0 to 8.)
- The observed common SD is 1.76, and inflated by 20%, the assumed SD is 1.76*1.2=2.11.
- Then the assumed effect size was 1.24/2.11=0.588.

Assuming normal distribution of the change in NPS, a common SD of 2.11, and a 25% dropout rate, with 120 patients per group, the study will have 98% power to detect the effect size of 0.588 using a two-sided test with alpha = 0.05 for the change in NPS at Week 24 in the dupilumab 300 mg q2w group.

Therefore, with a sample size of 120 patients per group, the combined power of the two coprimary efficacy endpoints is at least 93% for dupilumab 300 mg q2w group with alpha = 0.05 assuming no negative correlation between the two endpoints.

- The observed mean Lund-Mackay score (LMK) reduction of the dupilumab group with qw dosing in ACT12340 was 9.07 and the observed mean LMK reduction of the placebo group was 0.23. So the observed mean difference was 9.07-0.23=8.84. Discounting 20% for 300 q2w, the mean difference was assumed to be 8.84*0.8= 7.07 (LMK values range from 0 to 24.)
- The observed common SD is 4.58, and inflated by 20%, the assumed SD is 4.58*1.2=5.50.
- Then the assumed effect size was 7.07/5.50=1.285.
- Assuming normal distribution of the change in LMK, a common SD of 5.50, and a 25% dropout rate, with 120 patients per group, the study will have 99% power to detect the effect size of 1.285 using a two-sided test with alpha = 0.05 for the change in LMK at Week 24 in the dupilumab 300 mg q2w group.

With the same sample size of 120 patients per group, the combined power of the 3 co-primary efficacy endpoints for Japan is at least 92% for dupilumab 300 mg q2w group with alpha = 0.05 assuming no negative correlation between the 3 endpoints.

The sample size calculations were performed using nQuery Advisor 7.0.

1.4 STUDY PLAN

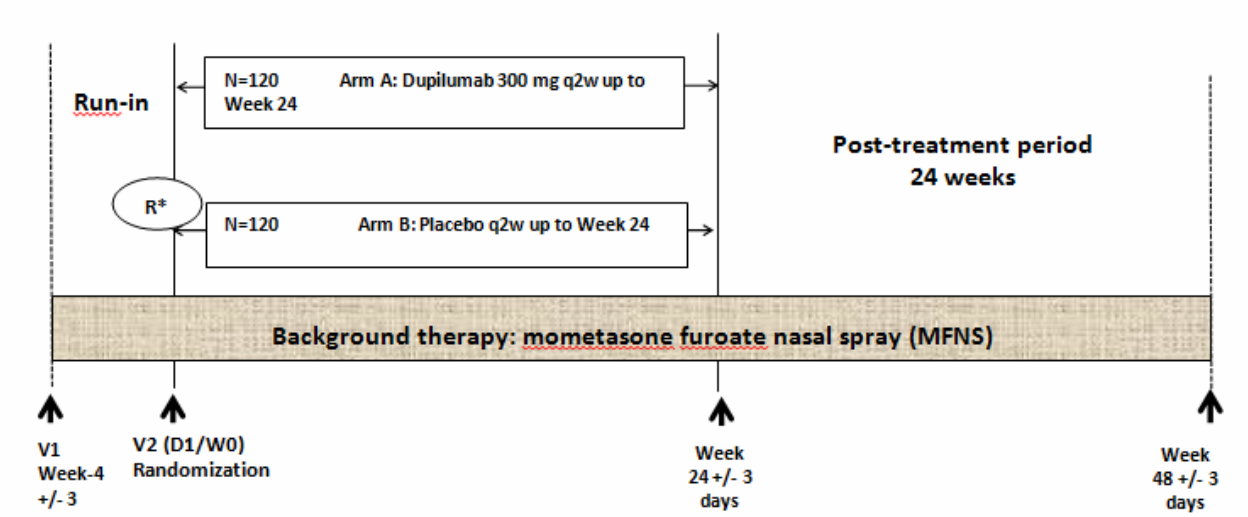
The clinical trial will consist of 3 periods:

- **Run-in period (4 weeks, +/-3 days):** to determine a patient's eligibility and for run-in/standardization of background INCS (MFNS) prior to randomization.
- **Randomized dupilumab/placebo treatment period (24 weeks +/-3 days):** to randomize the patient into a treatment arm and treat with dupilumab or placebo dose regimen.
- **Post-treatment period (24 weeks +/-3 days):** to continue to collect data for PK, immunogenicity, safety, and efficacy after the patient has completed the study drug treatment period.

If surgery is scheduled after the planned end of study, EOS visit will not be delayed. A follow up contact(s) should be performed around the time of planned surgery to document the surgery date and outcome. Surgery data will be collected until e-CRF closure of the trial.

1.4.1 Graphic study design

Please refer of Section 1.2 of the study protocol for the detailed study flow chart.



BID twice daily; D day; EOS end of study; EOT end of treatment; IMP investigational medicinal product; q2w every 2 weeks; MFNS mometasone furoate nasal spray; R* randomization; QD once daily; SC subcutaneous; V visit; W week

IMP: Regardless of the treatment group, all randomized patients will receive q2w SC administrations of dupilumab or placebo. Every other week IMP administrations must be separated by at least 11 days. At V2 the Investigator or delegate will perform the injection. After V2, every other week administration of IMP will be performed at the investigational site up to at least Week 8 (Visit 6). Patients will be monitored at the study site for at least 30 minutes or the minimum time required by your local regulator after injections. From Week 10, every other week home administration of IMP (patient, caregiver, or health care professional) is possible if the patient (or the caregiver) has been trained. If the patient (or caregiver) is unable or unwilling to administer IMP, arrangements must be made for qualified site personnel and/or healthcare professionals to administer IMP for the doses not scheduled to be given at the study site.

Non-investigational medicinal product: MFNS will be self-administered by the patient BID or QD (if they cannot tolerate BID). At each visit the Investigator must ensure that the patient has the necessary doses up to the next visit, knowing that one MFNS device (1 bottle) contains sufficient doses for: either 2 weeks of BID treatment/regimen or 4 weeks of QD treatment/regimen.

1.5 MODIFICATIONS TO THE STATISTICAL SECTION OF THE PROTOCOL

There is no change from the statistical section of the protocol.

1.6 STATISTICAL MODIFICATIONS MADE IN THE STATISTICAL ANALYSIS PLAN

NA

2 STATISTICAL AND ANALYTICAL PROCEDURES

2.1 ANALYSIS ENDPOINTS

2.1.1 Demographic and baseline characteristics

The baseline value of efficacy parameters is defined as the last available value up to randomization but prior to the first dose of study medication unless otherwise specified. The baseline value of safety parameters is defined as the last available value prior to the first dose of investigational medicinal product (IMP). The baseline value of the other parameters is defined as the last available value prior to the first dose of IMP if the patient is treated, or the last available value up to randomization if the patient is not exposed to IMP.

The calculation of the baseline for the daily symptom scores and the nasal peak inspiratory flow (NPIF) will be based on measurements obtained during the 7 days prior to randomization. For daily assessed electronic diary/NPIF parameters,

- If there are 4 or more measurements collected within 7 days prior to randomization, the baseline will be the average of these measurements.
- If less than 4 measurements are collected within 7 days prior to randomization, the baseline will be the average of the most recent 4 measurements prior to randomization (between V1 and V2).

All baseline safety and efficacy parameters (apart from those listed below) will be presented along with the on-treatment summary statistics in the safety and efficacy sections (Section 2.4.5 and Section 2.4.4).

Demographic characteristics

Demographic variables are

- Gender (Male, Female)
- Race (Caucasian/white, Black/of American descent, Asian/Oriental, American Indian or Alaska Native, Native Hawaiian or other Pacific Island, Multiple, Unknown)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Age in years (quantitative and qualitative variable: 18-64, 65–74, 75-84, and ≥85 years)
- Region (East Europe: Russia, Ukraine, Cz. Republic, Hungary, Poland, Romania, Bulgaria; Western Countries: France, Germany, Italy, United Kingdom, Netherlands, USA)

- Territory (North America: USA; European Union: Cz. Republic, Hungary, Poland, Romania, Bulgaria, France, Germany, Italy, United Kingdom and Netherlands; Rest of World: Russia and Ukraine)
- Weight in kg (quantitative and qualitative variable : <50, 50-<100 and ≥100 kg)
- BMI in kg/m^2 (quantitative and qualitative variable: <30, ≥30 kg/m^2)

Medical or surgical history

Medical (or surgical) history includes all the relevant medical (or surgical) history during the lifetime of the patient.

This information will be coded using the version of Medical Dictionary for Regulatory Activities (MedDRA) currently in effect at Sanofi at the time of database lock.

Comorbidity will be summarized separately. The following comorbid diseases will be summarized from eCRF pages which were filled in by investigators based on patient reporting. Asthma will be captured under disease characteristics at baseline.

Atopic dermatitis history (Yes, Ongoing condition)

Allergic conjunctivitis history (Yes, Ongoing condition)

Allergic rhinitis seasonal history (Yes, Ongoing condition)

Allergic rhinitis perennial history (Yes, Ongoing condition)

Eosinophilic esophagitis history (Yes, Ongoing condition)

Hives history (Yes, Ongoing condition)

Food allergy history (Yes, Ongoing condition)

Epsitaxis history (Yes, Ongoing condition)

NSAID exacerbated respiratory disease history (Yes, Ongoing condition)

Disease characteristics at baseline

The following baseline disease characteristics will be summarized by treatment group:

- NPS
- Lund-Mackay score
- AM symptom scores for nasal congestion/obstruction, decreased/loss of sense of smell, rhinorrhea (anterior/posterior nasal discharge) and total symptom score (TSS) respectively: For the daily symptoms the presence and the severity at 4 weeks and 8 weeks prior to screening will be provided.
- UPSIT smell test score
- SNOT-22
- Time since first diagnosis of nasal polyposis (years) to be derived as

  - (Year of randomization – Year of first diagnosis of nasal polyposis) + (month of randomization – month of first diagnosis of nasal polyposis)/12
- Age of onset of nasal polyposis (years)
- Number of previous surgeries for nasal polyposis (0, 1, 2, ≥3)
- Number of previous surgery for nasal polyposis by type:
  - Nasal/sinus endoscopy, surgical, with ethmoidectomy, total (anterior and posterior)
  - Nasal/sinus endoscopy, surgical, with maxillary and antrostomy with removal of tissue from maxillary sinus
  - Nasal/sinus endoscopy, surgical, with frontal sinus exploration, with or without removal of tissue from frontal sinus
  - Nasal/sinus endoscopy, surgical, with maxillary antrostomy
  - Nasal/sinus endoscopy, surgical, with sphenoidotomy with removal of tissue from the sphenoid sinus
  - Nasal/sinus endoscopy, surgical, with sphenoidotomy
  - Nasal/sinus endoscopy, surgical, with ethmoidectomy, partial (anterior)
  - Excision, nasal polyp(s), extensive
  - Excision, nasal polyp(s), simple
  - Excision or destruction (eg,laser), intranasal lesion; internal approach
  - Other
- Time since most recent nasal polyposis surgery (years) to be derived as
  - (Year of randomization – Year of most recent nasal surgery) + (month of randomization – month of most recent nasal surgery)/12
- Smoking history (former, current, never)
- Cessation prior to screening (months) for former smokers to be derived as
  - (Year of randomization – Year of cessation)×12 + (month of randomization – month of cessation)
  - Smoking quantity (cigarettes per day)
- Frequency of alcohol drinking in the past 12 months (never, occasional, at least monthly, at least weekly, at least daily)
  - For daily drinkers, provide the following
    - Number of standard alcohol drinks on a typical day (1 or 2, >2)
- Asthma comorbidity status (Yes, No)
  - For asthma patients, provide the following:
    - Time since first diagnosis of asthma (years) to be derived as

(Year of randomization – Year of first diagnosis of asthma) + (month of randomization – month of first diagnosis of asthma)/12

- Time since last asthma exacerbation (days) to be derived as

  Date of randomization – Date of last asthma exacerbation

- Number of asthma exacerbations experienced 1 year before visit 1 (comorbid asthma patients only, quantitative variable and qualitative variable: 0,1,2,3, ≥4)
- Age of onset of asthma

- Number of courses of systemic corticosteroid (SCS) use during the past 2 years (0, 1, 2, 3, 4, ≥ 5)
  - A course of SCS is considered continuous if treatment is separated by less than 7 days.
- Number of days of systemic corticosteroid (SCS) use during the past 2 years (0, >0 - ≤7, >7 - ≤14, >14 - ≤ 21, >21 - ≤28, >28)

Any technical details related to computation, dates, and imputations for missing dates are described in Section 2.5.

2.1.2 Prior or concomitant medications

All medications taken within 30 days prior to screening and until end of the study, including nasal polyposis medications (except for MFNS), chronic rhinosinusitis medications (except for MFNS), asthma medications, and all other medications are to be reported in the case report form (CRF) pages.

All medications will be coded using the World Health Organization-Drug Dictionary (WHO-DD) version currently in effect at Sanofi at the time of database lock.

Prior medications are those the patient used prior to first IMP injection. Prior medications can be discontinued before first administration or can be ongoing during treatment phase.

Concomitant medications are any treatments received by the patient concomitantly to the IMP, anytime from the first administration of IMP to the last administration of IMP + 98 days. A given medication can be classified as a prior medication, concomitant medication and post-treatment medication at the same time.

Post-treatment medications are those the patient took in the period from the last administration of IMP + 99 days to the end of the study.

Any technical details related to computation, dates, imputation for missing dates are described in Section 2.5.

2.1.2.1 MFNS controller medication

On a daily basis throughout the study, mometasone furoate nasal spray (MFNS) will be self-administered (and recorded using an e-diary) by the patients. Starting from V1, they will administer two actuations (50 μg/actuation) in each nostril BID (total daily dose of 400 μg), unless they are unable to tolerate the BID regimen or this dose is not approved in specific countries, in which case, they will follow a once daily (QD) regimen (total daily dose of 200 μg).

2.1.3 Efficacy endpoints

Baseline for efficacy endpoints is defined as the last non-missing value up to randomization but prior to the administration of the first IMP unless otherwise specified.

2.1.3.1 Primary efficacy endpoint(s)

There are 2 co-primary endpoints for this study for all countries except for Japan:

1. **Change from baseline in the nasal congestion/obstruction (NC) score at Week 24:** The NC score is assessed by the patient on a daily basis from V1 and throughout the study, using an e-diary, on a 0-3 categorical scale (1):

Nasal congestion/obstruction will be scored as a reflective score (evaluation of symptom severity over the past 24 hours) by the patient:

Scale	Symptoms
0	No symptoms
1	Mild symptoms (symptoms clearly present, but minimal awareness and easily tolerated)
2	Moderate symptoms (definite awareness of symptoms that is bothersome but tolerable)
3	Severe symptoms (symptoms that are hard to tolerate, cause interference with activities or daily living)

The e-diary is dispensed at V1 and information is downloaded from this device on the other protocol specified visit days.

A severity ≥2 on V1 and a weekly average severity greater than 1 at time of randomization (V2) is required for patient eligibility.

For baseline calculation (used to determine eligibility and to assess efficacy),

- If there are 4 or more measurements collected within 7 days prior to randomization available, the baseline will be the average of these measurements.

- If less than 4 measurements are collected within 7 days prior to randomization, the baseline will be the average of the most recent 4 measurements prior to randomization (between V1 and V2).

For the baseline to EOT analysis, the average of all scores from the preceding 4 weeks to the corresponding visit will be used as the analysis score for that visit.

2. **Change from baseline in the NPS at Week 24:** The NPS (2)(3)(4) is assessed by centralized scoring of nasal endoscopy video recordings. The score (NPS) is the sum of the right and left nostril scores, as evaluated by means of nasal endoscopy. NP is graded based on polyp size described below.

Polyp Score	Polyp Size
0	No polyps
1	Small polyps in the middle meatus not reaching below the inferior border of the middle turbinate
2	Polyps reaching below the lower border of the middle turbinate
3	Large polyps reaching the lower border of the inferior turbinate or polyps medial to the middle turbinate
4	Large polyps causing complete obstruction of the inferior nasal cavity

Nasal endoscopy should be performed before the administration of IMP and preceded by local administration of anesthetic drugs in combination with a decongestant.

Standard video sequences of the nasal endoscopy will be sent to centralized reader for blinded review. The sites will remove patient-identifying information from the imaging data prior to sending the imaging data to the central reader. Centralized imaging data assessments and scoring by independent physician reviewer(s) for the imaging data will be performed for all endoscopies. Specifically, a double central reading with adjudication process will be performed to improve scoring quality.

To confirm eligibility at V2, only the V1 central reading will be made available to the site. In addition, at V2 the investigator will perform the nasal endoscopy to confirm eligibility score and enter the result in the e-CRF. Thus the patient is considered eligible based on a V1 central reading followed by a V2 local reading NPS total score of 5 or more and a score of at least 2 in each side. The final results of central reading from Visit 2 onward will be made available after the study.

The adjudication for V1 will not be required if both readers score ineligible (total NPS score less than 5 or at least one side less than 2, in which case the patient will be considered as ineligible) or both readers score eligible (total NPS score of 5 or more and at least 2 in each side, in which case the patient will be considered as eligible). Unreadable scores (bad quality videos) at V1 even from one reader should have endoscopy redone as soon as possible (requiring patient to come back at site). In the case the two readers differ on eligibility for the study, an adjudicator will read the endoscopy again independently. The total NPS score and each side score to determine eligibility will be based on the average of the adjudicator and one of the two readers whose total NPS score

is closer to the adjudicator's reading (in the case that the two readers' scores are different but equidistant from the adjudicator's score, the reader with the higher score will be used in the calculation). In the special case that the two readers give different eligibility opinions but the total NPS scores are the same (the only possibility of the total NPS score for such a case is 5), eligibility will follow the adjudicator's opinion.

For V2 and onward (excluding V1), if the total NPS scores assigned by two independent readers differ less than or equal to 1, the total NPS score for analysis for this endoscopy will be the average of the two readings; if they differ more than 1, an adjudicator will read the endoscopy again independently, and the total NPS score for analysis will be the average of the total NPS scores by the adjudicator and one of the two readers whose total NPS score is closer to the adjudicator's reading (in the case that the two readers' total NPS scores are equidistant from the adjudicator's total NPS score, the higher score will be used in the calculation). If the endoscopy is scored unreadable by both readers, the analysis score will be set to missing. If one and only one of the two readers score unreadable, the endoscopy will be read again independently by the adjudicator. Then the average of the two available readings (if the adjudicator gives a readable score) or missing value (if the adjudicator scores it as unreadable) will be set as the total NPS score for analysis.

More details regarding the study eligibility determination based on the double central reading with adjudication are listed in Appendix B.

For the analysis of primary endpoint, central reading of V2 will be used as the baseline. In the scenario that V2 central reading scores are missing, V1 central reading will be used instead and adjudication will be required if the total NPS scores by the two readers at V1 differ more than 1 and the adjudication has not been performed. Calculations of the analysis score for such V1 readings will follow the aforementioned approach for V2 and onward.

3. Change from baseline in Lund-Mackay score (LMK) at Week 24

For Japan, in addition to the two co-primary endpoints above, Lund-Mackay score (LMK) will also be a co-primary endpoint (for countries other than Japan, change from baseline in LMK at Week 24 will be a key secondary endpoint) which is defined as follows:

The LMK system is based on localization with points given for degree of opacification: 0 = normal, 1 = partial opacification, 2 = total opacification. These points are then applied to the maxillary, anterior ethmoid, posterior ethmoid, sphenoid and frontal sinuses on each side. The Osteomeatal complex (OC) is graded as 0 = not occluded, or 2 = occluded. Adding the scores of these 6 parameters will derive a maximum score of 12 per side (left and right). The total score of the two sides will be used for analysis. This scoring system has been validated in several studies (5)(6)(7).

For patients in whom the OC is missing (because of a previous surgery) the reader should consider the location of the former OC and provide a scoring (as if the OC was there).

One reviewer will review CT images for all subjects enrolled in the clinical trial. The reviewer will perform the independent review under guidance of the study protocol and GCP.

A CT scan should be performed anytime during the run-in period before first administration of IMP, and at V8 (Week 24). CT scan is optional at V10 (Week 48). Whenever possible a cone beam CT scan should be utilized. In countries for which a specific approval procedure for the CT scan is required by a different committee than the local independent ethics committee (IEC)/institutional review board (IRB), patients may be enrolled using a CT available in the previous year or perform an MRI of the sinuses between V1 and V2. These countries are exempted from all the planned study CT scans until approval from these committees is received.

2.1.3.2 Secondary efficacy endpoint(s)

2.1.3.2.1 Key secondary endpoint

2.1.3.2.1.1 Lund-Mackay score

Lund-Mackay score is a secondary efficacy endpoint in countries other than Japan. Details of this endpoint are presented in Section 2.1.3.1.

2.1.3.2.1.2 Disease specific daily symptom assessment and total symptom score (TSS)

On a daily basis from V1 and throughout the study, the patient will use an e-diary to:

- Respond to the morning individual rhinosinusitis symptom questions using a 0 to 3 categorical scale (where 0 = no symptoms, 1 = mild symptoms, 2 = moderate symptoms and 3 = severe symptoms) :
  - Congestion and/or obstruction.
  - Loss of sense of smell.
  - Anterior rhinorrhea (runny nose).
  - Posterior rhinorrhea (postnasal drip).

The TSS is a composite score (ranging between 0 and 9) consisting of the sum of the following symptoms assessed daily in the morning: nasal congestion/obstruction, decreased/loss of sense of smell, rhinorrhea (average of anterior/posterior nasal discharge).

The symptom scores at baseline and from baseline to EOT analysis will be calculated using the same approach as for NC as described in Section 2.1.1 and Section 2.1.3.1. If there are 4 or more measurements collected within 7 days prior to randomization, the baseline will be the average of these measurements; if less than 4 measurements are collected, the baseline will be the average of the most recent 4 prior to randomization. For the baseline to EOT analysis, 4 weeks average of the symptom scores prior to the corresponding visit will be used as the analysis score for that visit.

2.1.3.2.1.3 Smell test: University of Pennsylvania Smell Identification Test (UPSIT)

The UPSIT (UPSIT 40 odorant test) is a rapid and easy-to-administer method to quantitatively assess human olfactory function. The UPSIT shows a high test-retest reliability (r: 0.981) and

scores on this test are strongly correlated with the detection threshold for phenyl ethyl alcohol in the same individuals. When the UPSIT is administered in the standardized manner, clinical subjects show a high degree of uniformity in UPSIT performance when tested in different laboratories.

The test consists of 4 booklets, each containing 10 odorants with one odorant per page. Above each odorant strip is a multiple-choice question with four alternative words to describe the odor. The subject is asked to release the odorant by rubbing the brown-strip with the tip of a pencil and to indicate which of four words best describes the odor. Thus each subject receives a score out of 40 possible correct answers. The final score is recorded in the e-CRF. The odorants of the UPSIT test utilized in this study take into account cultural differences.

2.1.3.2.1.4 Decreased/loss of sense of smell

The decreased/loss of sense of smell severity is assessed by the patient on a daily basis from V1 and throughout the study, using an e-diary to using a 0-3 categorical scale (where 0 = no symptoms, 1 = mild symptoms, 2 = moderate symptoms and 3 = severe symptoms). The symptom scores at baseline and from baseline to EOT analysis will be calculated using the same approach as for NC.

2.1.3.2.1.5 22-Item sino-nasal outcome test (SNOT-22)

The SNOT-22 is a validated questionnaire to assess the impact of chronic rhinosinusitis on HRQoL (Appendix D). The SNOT-22 has 22 items on a 5-category scale applicable to sino-nasal conditions and surgical treatments. The range of the global score is 0-110 with a minimal important difference (MID) of 8.9 (8). The MID is the smallest mean change from baseline (point estimates) that will be interpreted as important; lower scores indicate less impact. The recall period is the prior 2 weeks. At baseline, the patient is also asked to rank the 5 most important items affecting their health.

The global score is the sum of response to each of the 22 questions. For responses with some item scores are missing, the global score will be imputed as mean of the completed scores× 22, providing at least 50% of items have been completed. If more than 50% of items are missing, the global score is set to missing.

2.1.3.2.1.6 Proportion of patients requiring rescue treatment defined as: use systemic corticosteroids or NP surgery (actual or planned) during the treatment period

SCS for any reason

Systemic corticosteroids for rescue treatment of nasal polyps or for another reason administered during study treatment are recorded by the Investigator (or designee) who records the date and dosing information (daily dose, duration, INN) on the appropriate page(s) of the e-CRF. Indication for SCS use will be also captured by selecting one or more of the following categories:

1. Nasal polyposis.

2. Asthma.
3. Other respiratory disease (specify).
4. Other ear, nose or throat disease (specify).
5. Other reason (specify).

Surgery (actual or planned) for NP

For patients who undergo or are planned for sino-nasal surgery for NP, the reason (worsening signs and/or symptoms during the study), the date of decision for surgery, the planned date and the effective surgery date, the type and outcome of surgery will be recorded in a specific e-CRF page.

If surgery is performed during the study treatment period, patient and investigator may decide to continue IMP up to the time of surgery or EOT whichever date comes first. At the time of surgery patients will be permanently discontinued from study treatment and assessed as soon as possible using the procedures normally planned for the EOT Visit and will be instructed to return to the study site as described in the protocol. An AE or serious adverse event (SAE) page will be completed.

If surgery is performed during the follow-up, the patients will be assessed according to the procedures normally planned for the EOS Visit and will be instructed to return to the study site as described in protocol. An AE or SAE page will be completed.

If surgery date is scheduled after the planned end of study, EOS visit will not be delayed. A follow up contact(s) should be performed around the time of planned surgery to document the surgery date and outcome. Surgery data will be collected until e-CRF closure of the trial.

2.1.3.2.1.7 Spirometry

Spirometry was performed at local level (study site or another facility) in the morning after withholding the last dose of salbutamol/albuterol or levosalbutamol/levalbuterol for at least 6 hours. Forced expiratory volume in one second (FEV1), forced vital capacity (FVC) and forced expiratory flow at 25% to 75% of forced vital capacity (FEF 25-75) were determined at the designated treatment visits (9). The results of FEV1 (score in L, % of predicted normal,), FVC and FEF 25-75 were recorded in the e-CRF anytime during run-in period (before V2) for all patients and in patients with asthma for the other scheduled visits during the randomized treatment period.

Whenever possible, the same spirometer and standard spirometric techniques, including calibration, should be used to perform spirometry at all visits and, the same person should perform the measurements.

2.1.3.2.2 Other secondary efficacy endpoints

2.1.3.2.2.1 Visual analogue scale

The visual analogue scale (VAS) for rhinosinusitis is used to evaluate the total severity (1). Rhinosinusitis disease can be divided into MILD, MODERATE and SEVERE based on total severity VAS score (0 to 10 cm):

- MILD = VAS 0 to 3.
- MODERATE = VAS >3 to 7.
- SEVERE = VAS >7 to 10.

The patient is asked to indicate on a VAS the answer to the question below:

"How troublesome are your symptoms of your rhinosinusitis"

The VAS ranks from 0 (Not troublesome) to 10 (Worst thinkable troublesome).

2.1.3.2.2.2 Nasal peak inspiratory flow

Nasal peak inspiratory flow (NPIF) evaluation represents a physiologic measure of the air flow through both nasal cavities during forced inspiration expressed in liters per minute. The NPIF is the best validated technique for the evaluation of nasal flow through the nose. Nasal inspiration correlates most with the subjective feeling of obstruction and is the best validated technique for monitoring nasal flow in clinical trials.

On V1 patients will be issued an NPIF meter for recording morning (AM) NPIF. Patients will be instructed on the use of the device, and written instructions on the use of the NPIF meter will be provided to the patients. In addition, the Investigator will instruct the patients on how to record the following variables in the e-diary on a daily basis.

- AM NPIF performed within 15 minutes after arising (before 12 noon) prior to taking MFNS.

Three NPIF efforts will be performed by the patient; all 3 values will be recorded by the patient in the e-diary, and the highest value will be used for evaluation. The procedure takes about 5 minutes.

Baseline AM NPIF will be the mean AM measurement recorded for the 7 days prior to the first dose of IMP. If less than 4 measurements are collected within 7 days prior to randomization, the average of the most recent 4 prior to randomization during the run-in period will serve as the baseline as described in Section 2.1.1.

The NPIF will be performed daily from V1 to V8 (Week 24). After V8 the NPIF will be performed every 4 weeks.

The nasal flow is expressed in liter per minute, and consecutive measurements are performed.

2.1.3.2.2.3 Asthma Control Questionnaire, 6-question version (ACQ-6) in those patients comorbid with asthma

The asthma control questionnaire-6 (ACQ-6) was designed to measure both the adequacy of asthma control and change in asthma control which occurs either spontaneously or as a result of treatment. Only patients with co-morbid asthma will be asked to complete the questionnaire in the e-diary during clinic visits. Patients should complete the questionnaire before the spirometry test.

The asthma control questionnaire-6 (ACQ-6) has 6 questions which assess the most common asthma symptoms:

- Frequency in past week awoken by asthma during the night.
- Severity of asthma symptoms in the morning.
- Limitation of daily activities due to asthma.
- Shortness of breath due to asthma.
- Frequency of wheezing, and
- Short-acting bronchodilator use.

Patients are asked to recall how their asthma has been during the previous week and to respond to the symptom questions on a 7-point scale (0 = no impairment, 6 = maximum impairment) (see Appendix C).

A global score is calculated: the questions are equally weighted and the ACQ-6 score is the mean of the 6 questions and, therefore, between 0 (totally controlled) and 6 (severely uncontrolled). Higher score indicates lower asthma control. Patients with a score below 1.0 reflect adequately controlled asthma and patients with scores above 1.0 reflect inadequately controlled asthma. On the 7-point scale of the ACQ-6, a change or difference in score of 0.5 is the smallest change that can be considered clinically important, corresponding to the Minimal Clinically Important Difference (MCID) defined by the developer.

Measurement properties such as reliability and ability to detect change have been documented in the literature (10).

2.1.4 Safety endpoints

The safety analysis will be based on the reported adverse events and other safety information, such as clinical laboratory data, vital signs, ECG and physical examination.

Observation period

The observation period will be divided into 4 epochs:

- The **screening** epoch is defined as the time from the signed informed consent date up to the time prior to first administration of the IMP.

- The **treatment** epoch is defined as the time from the first administration of the IMP to the last administration of the IMP + 14 days.
- The **residual treatment** epoch is defined as the time from the last administration of the IMP +15 days to the last administration of the IMP + 98 days.

The treatment-emergent adverse event period will include both **treatment** and **residual treatment** epochs.

- The **post-treatment** epoch is defined as the period of time starting the day after the end of the treatment-emergent adverse event period up to the end of the study (defined as last protocol planned visit or the resolution/stabilization of all serious adverse events and adverse events with pre-specified monitoring).

The on-study observation period is defined as the time from start of treatment until the end of the study (defined as last protocol planned visit or lost to follow-up or the resolution/stabilization of all serious adverse events and adverse events with pre-specified monitoring).

2.1.4.1 Adverse events variables

Adverse event observation period

- Pretreatment adverse events are adverse events that developed or worsened or became serious during the screening epoch.
- Treatment-emergent adverse events are adverse events that developed or worsened or became serious during the treatment-emergent adverse event period
- Post-treatment adverse events are adverse events that developed or worsened or became serious during the post-treatment epoch.

All adverse events (including serious adverse events and adverse events with pre-specified monitoring/of special interests) will be coded to a lower-level term (LLT), preferred term (PT), high-level term (HLT), high-level group term (HLGT), and associated primary system organ class (SOC) using the version of Medical Dictionary for Regulatory Activities (MedDRA) currently in effect at Sanofi at the time of database lock.

The occurrence of adverse events (including serious adverse events and adverse events with pre-specified monitoring) will be recorded from the time of signed informed consent until the end of the study.

Adverse events of special interests (AESI) and other selected AE groupings will be searched based on the criteria in Table 1.

Table 1 Criteria for adverse events of special interest and other selected AE groupings

AE Grouping	Criteria
AESI	
Anaphylactic reaction	Anaphylactic reaction algorithmic approach (*Introductory Guide for Standardised MedDRA Queries (SMQs) Version 18.1*): includes anaphylactic reaction narrow SMQ (20000021) terms; for selection based on occurrence of multiple symptoms, the symptoms must have occurred within 24 hours of each other
Hypersensitivity (medically reviewed)	SMQ hypersensitivity (20000214) narrow search) and [AE corrective treatment/therapy='Y' or Action taken with IMP='Drug withdrawn' or Action taken with IMP='Drug interrupted'] followed by blinded medical review (documented process) for selection of relevant systemic hypersensitivity events
Serious injection site reactions or severe injection site reactions that last longer than 24 hours	HLT = 'Injection site reaction' and either with serious status, or with severe status and (AE end date/time - AE start date/time) ≥ 24 hours or ongoing
Severe or serious infection	Primary SOC = 'Infections and infestations' and (Intensity='Severe' or Serious='Y')
Parasitic infection	The Infection Type 'Parasitic' was checked on eCRF page "Infection Defined as AESI Complementary Form"
Opportunistic infection	The Infection Type 'Opportunistic' was checked on eCRF page "Infection Defined as AESI Complementary Form"
Drug-related hepatic disorder	Drug-related hepatic disorders-Comprehensive search narrow SMQ (20000006)
Pregnancy	Primary SOC 'Pregnancy, puerperium and perinatal conditions' or PT in (Aborted pregnancy, False negative pregnancy test, Pregnancy test positive, Pregnancy test urine positive, Ectopic pregnancy termination)
Symptomatic overdose with IMP	The question "Is the event a Symptomatic Overdose with IMP?" is answered "Yes" on the eCRF page "Adverse Event".
Symptomatic overdose with NIMP	The question "Is the event a Symptomatic Overdose with NIMP?" is answered "Yes" on the eCRF page "Adverse Event".
Other selected AE grouping	
Injection site reaction	HLT = 'Injection site reaction'
Malignancy	Sub-SMQ (20000091)– Malignant or unspecified tumors
Suicidal behavior	PT in (Completed suicide, Suicidal ideation, Depression suicidal, Suicidal behavior, Suicide attempt)
Partner pregnancy	PT in (Pregnancy of partner, Miscarriage of partner)
Epistaxis/nose bleeding	PT in (Epistaxis, Nasal septum haematoma)
Conjunctivitis (narrow company medical query)	PT in (Conjunctivitis, Conjunctivitis allergic, Conjunctivitis bacterial, Conjunctivitis viral, Atopic keratoconjunctivitis)

AE Grouping	Criteria
Conjunctivitis (broad company medical query)	PT in (Conjunctivitis, Conjunctivitis allergic, Conjunctivitis bacterial, Conjunctivitis viral, Atopic keratoconjunctivitis, Blepharitis, Dry eye, Eye irritation, Eye pruritus, Lacrimation increased, Eye discharge, Foreign body sensation in eyes, Photophobia, Xerophthalmia, Ocular hyperaemia, Conjunctival hyperaemia)
Hypereosinophilia	HLT = 'Eosinophilic disorders' or PT = 'Eosinophil count increased'

2.1.4.2 Deaths

The deaths observation period are per the observation periods defined above.

- Death on-study: deaths occurring during the on-study observation period
- Death on-treatment: deaths occurring during the treatment-emergent adverse event period
- Death post-study: deaths occurring after the end of the study

2.1.4.3 Laboratory safety variables

Clinical laboratory data consists of blood analysis, including hematology, clinical chemistry, and urinalysis. Clinical laboratory values after conversion will be analyzed into standard international units and international units will be used in all listings and tables.

Blood samples for clinical laboratories will be taken at Visit 1(Week -4±1), Visit 2 (Week 0), Visit 7 (Week 16), Visit 8 (Week 24), and early termination unless otherwise specified. The laboratory parameters will be classified as follows:

- Hematology
  - **Red blood cells and platelets and coagulation**: hemoglobin, hematocrit, red blood cell count, platelet count
  - **White blood cells:** white blood cell count, neutrophils, lymphocytes, monocytes, basophils, eosinophils
- Clinical chemistry
  - **Metabolism:** glucose, total cholesterol, total protein, creatine phosphokinase.
  - **Electrolytes**: sodium, potassium, chloride, bicarbonate.
  - **Renal function**: creatinine, blood urea nitrogen, uric acid.
  - **Liver function**: alanine aminotransferase, aspartate aminotransferase, alkaline phosphatase, lactate dehydrogenase, total bilirubin (in case of values above the normal range, differentiation in conjugated and non-conjugated bilirubin), albumin.
  - **Pregnancy test**: Serum β-human chorionic gonadotropin (all female patients) will be performed at screening (V1) in women of childbearing potential, and a urine dipstick pregnancy test will be performed at V2 prior to randomization and every 4 weeks thereafter. A negative result must be obtained at V1 and V2 prior to randomization. In

case of positive urinary test the study treatment will be withheld and a serum pregnancy test to confirm the pregnancy should be performed as soon as possible. Pregnancy will lead to definitive treatment discontinuation in all cases.

- **Hepatitis screen:** hepatitis B surface antigen (HBs Ag), hepatitis B Surface antibody (HBs Ab), hepatitis B core antibody (HBc Ab), hepatitis C virus antibodies (HCV Ab), In case of results showing HBs Ag (negative), HBs Ab (negative) and HBc Ab (positive), an HBV DNA testing may be performed prior to randomization to rule out a false positivity if the investigator believes the patient is a false positive, or to clarify the serological status if the investigator finds it unclear to interpret in absence of known HBV infection. In case of results showing HCV Ab (positive), an HCV RNA testing may be performed to rule out a false positivity, if the investigator believes the patient is a false positive.
- **Hepatitis B viral load** will be tested at V2, and V8 for patients in Japan or other countries/regions if there is local regulatory requirement who are HBs Ag negative and HBs Ab positive at screening.
- **HIV screen**: Anti-HIV-1 and HIV-2 antibodies will be tested at Visit 1.
- **Anti-nuclear antibody (ANA)** will be tested at Visit 1.
- Note: Anti-ds DNA antibody will be tested if ANA is positive (≥1:160 titer).

Technical formulas are described in Section 2.5.1.

2.1.4.4 Vital signs variables

Vital signs, including blood pressure (mmHg), heart rate (beats per minute), respiratory rate (breaths per minute), body temperature (degrees Celsius) and body weight (kg) will be measured at the screening and randomization visits (Visits 1 and 2) and subsequent visits pre-specified in the flowchart. Height (cm) will be measured at screening (Visit 1) only. Vital signs will be measured in the sitting position using the same arm (preferably) at each visit, and will be measured prior to receiving IMP at the clinic visits.

2.1.4.5 Electrocardiogram variables

A standard 12-lead electrocardiogram (ECG) will be performed at the sites at the time points noted in the Study Flowchart (Section 1.4.1) to monitor any potential abnormality. In case of an abnormal ECG finding, the Investigator should enter details into the e-CRF. At V2, the investigator should use their medical judgment to consider whether the patient is eligible for the study.

2.1.4.6 Physical Examination

Physical examinations will be performed at Visit 1(Week -4±1), Visit 8 (Week 24), Visit 9 (Week 36), and Visit 10 (Week 48) including an assessment of general appearance, skin, eyes,

ear/nose/throat, heart, chest, abdomen, reflexes, lymph nodes, spine, and extremities. All deviations from normal will be recorded, including those attributable to the patient's disease.

2.1.5 Pharmacokinetic variables

Predose functional dupilumab concentrations in serum at V2 (Day 1), dupilumab trough concentrations at Week 4, Week 8, Week 16, Week 24 (EOT), and post-treatment at Week 36 and Week 48 (EOS) will be provided.

Anti-dupilumab antibody (ADA) status (negative or titer value, if positive in the ADA assay) at Visit 2 (Day 1), Week 8, Week 16, Week 24, Week 36 and Week 48 will be provided. Patients who discontinue early from treatment may be asked to return to the clinic to have additional ADA samples collected for analysis based on the overall clinical presentation at that time.

ADA incidence will be classified as the following:

Pre-existing immunoreactivity are defined as:

An ADA positive response in the assay at baseline with all post-treatment ADA results negative, OR an ADA positive response at baseline with all post-treatment ADA responses less than 4-fold over baseline titer levels.

Treatment-emergent anti-drug antibodies are defined as:

A positive response in the ADA assay post first dose, when baseline results are negative or missing.

Treatment-emergent ADA responses are further classified as Persistent, Indeterminate or Transient

A) Persistent Response- defined as a treatment-emergent ADA response with two or more consecutive ADA positive sampling time points, separated by greater than (>) 12-week period (84 days) ,with no ADA negative samples in between.
B) Indeterminate Response- defined as a treatment-emergent response with only the last collected sample positive in the ADA assay
C) Transient Response - defined as a treatment-emergent response that is not considered persistent OR indeterminate

Treatment-boosted response is defined as:

An ADA positive response in the assay post first dose that is greater-than or equal to 4-fold over baseline titer levels, when baseline results are positive.

Titer values (Titer value category)

- Low (Titer < 1000)
- Moderate (1,000 ≤ Titer ≤ 10,000)

- High (Titer > 10,000)

2.1.6 Pharmacodynamic [redacted] endpoints

Blood biomarkers (Eotaxin-3, Thymus and Activation-Regulated Chemokine [TARC] and periostin) will all be assayed at Visit 2(Week 0).

Serum Total IgE, allergen-specific IgE panel sampling including Staph enterotoxins IgE will be assayed at Visit 2(Week 0) and Visit 8(Week 24).

Urine biomarkers (Leukotriene E4 and a metabolite of prostaglandin D2 (PGDM) and creatinine) will be measured in morning spot urine samples using validated quantitative assays.

[redacted]

Using a nasal swab, mucus will be collected from the middle meatus of each nostril and assessed for relative abundance of microbes, in all patients prior to MFNS administration. Mucus will be assayed at Visit 2(Week 0) and Visit 8(Week 24).

2.1.7 Health Related Quality-of-life endpoints

2.1.7.1 Euro-QOL-5D

The European quality of life-5D scale (EQ-5D) (Appendix E) is a standardized HRQoL questionnaire developed by the EuroQol Group in order to provide a simple, generic measure of health for clinical and economic appraisal (11). EQ-5D is designed for self-completion by patients.

The EQ-5D essentially consists of 2 pages: the EQ-5D descriptive system and the EQ VAS. The EQ-5D comprises 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The EQ VAS records the respondent's self-rated health on a vertical VAS. The EQ VAS 'thermometer' has endpoints of 100 (best imaginable health state) at the top and 0 (worst imaginable health state) at the bottom. EQ-5D self-reported VAS data generates information on the self-perceived overall health-related quality of life.

EQ-5D will be assessed at Visit 2 (Week 0) and Visit 8 (Week 24).

2.1.7.2 Short Form 36 (SF-36) Version 2

The SF-36 (Short-Form-36 (SF-36)-Version 2.0) (Appendix F) is a generic questionnaire measuring general health status (QoL) in the last 4 weeks before completing the questionnaire.

The SF-36 is a 36-item questionnaire that measures eight multi-item dimensions of health: physical functioning (10 items) social functioning (2 items) role limitations due to physical problems (4 items), role limitations due to emotional problems (3 items), mental health (5 items), energy/vitality (4 items), pain (2 items), and general health perception (5 items).

For each dimension, item scores are coded, summed, and transformed on to a scale from 0 (worst possible health state measured by the questionnaire) to 100 (best possible health state). Two standardized summary scores can also be calculated from the SF-36; the physical component summary (PCS) and the mental health component summary (MCS).

In this study SF-36 will be assessed only at V2 to evaluate the baseline QoL impact in the defined phase 3 population.

2.1.8 Health economic endpoints

Health care resource utilization (reliever medication, specialist visit, hospitalization, emergency or urgent medical care facility visit), and productivity (sick leaves, school days loss, usual activity loss etc.) will be collected at Visit 2 (Week 0), Visit 6 (Week 8), Visit 7 (Week 16), and Visit 8 (Week 24) through eCRF.

2.2 DISPOSITION OF PATIENTS

This section describes patient disposition for both patient study status and the patient analysis populations.

Screened patients are defined as any patients who signed the informed consent.

Randomized patients consist of all patients with a signed informed consent form who have had a treatment kit number allocated and recorded in the IRT database, regardless of whether the treatment kit was used.

For patient study status, the total number of patients in each of the following categories will be presented in the clinical study report using a flowchart diagram or summary table:

- Screened patients
- Screen failure patients and reasons for screen failure
- Nonrandomized but treated patients
- Randomized patients
- Randomized but not treated patients
- Randomized and treated patients
- Patients who did not complete the 24-week treatment as per protocol
- Patients who did not complete the 48-week follow-up as per protocol

- Patients who withdraw from study
- Patients who withdraw from study by main reason for study discontinuation
- Patients who withdraw from study prior to Week 24
- Patients who withdraw from study prior to Week 24 by main reason for study discontinuation.
- Vital status at last study contact

For all categories of patients (except for the screened and nonrandomized categories) percentages will be calculated using the number of randomized patients as the denominator. Whether patients received rescue treatment (SCS or NP surgery) by Week 24 and Week 48 will be provided in tables. The tables will list patient numbers and percentages by treatment group and type of rescue treatment (SCS or NP surgery).

All critical or major deviations potentially impacting efficacy analyses, randomization, and drug-dispensing irregularities, and other major or critical deviations will be summarized in tables giving numbers and percentages of deviations by treatment group.

Additionally, efficacy populations will be summarized by number of patients on the randomized population.

- Efficacy population: intent-to-treat (ITT) population

The analysis populations for safety, pharmacokinetics/pharmacodynamics will be summarized by number of patents on the safety population.

- Safety population
- Systemic drug concentration population
- PK population
- ADA population

2.2.1 Randomization and drug dispensing irregularities

Randomization and drug-dispensing irregularities occur whenever:

1. A randomization is not in accordance with the protocol-defined randomization method, such as a) an ineligible patient is randomized, b) a patient is randomized based on an incorrect stratum, c) a patient is randomized twice, or d) in a dynamic randomization scheme the treatment assignment is, in fact, not random, due to a computer program error.

   OR

2. A patient is dispensed an IMP kit not allocated by the protocol-defined randomization, such as a) a patient at any time in the study is dispensed a different treatment kit than as randomized (which may or may not contain the correct-as-randomized IMP), or b) a nonrandomized patient is treated with IMP reserved for randomized patients.

Randomization and drug-dispensing irregularities will be monitored throughout the study and reviewed on an ongoing basis.

All randomization and drug-dispensing irregularities will be documented in the clinical study report. If the number of irregularities is large enough to make a tabular summary useful, the irregularities will be categorized and summarized among randomized patients (number and percentages). Nonrandomized, treated patients will be described separately.

Randomization and drug-dispensing irregularities to be prospectively identified include but are not limited to:

Randomization and drug allocation irregularities
Kit dispensation without IRT transaction
Erroneous kit dispensation
Kit not available
Randomization by error
Patient randomized twice
Forced randomization
Stratification error
Patient switched to another site

2.3 ANALYSIS POPULATIONS

Patients treated without being randomized will not be considered randomized and will not be included in any efficacy population.

The randomized population includes any patient who has been allocated to a randomized treatment regardless of whether the treatment kit was used.

For any patient randomized more than once, only the data associated with the first randomization will be used in any analysis population. The safety experience associated with any later randomization will be assessed separately.

2.3.1 Efficacy populations

The primary efficacy analysis population will be the ITT population as defined in Section 2.3.1.1.

2.3.1.1 Intent–to-treat population

The intent-to-treat population is the randomized population analyzed according to the treatment group allocated by randomization.

Patients will be analyzed in the treatment group to which they are randomized.

2.3.2 Safety population

The safety population is defined as: All patients who actually received at least 1 dose or part of a dose of the IMP, analyzed according to the treatment patients actually received.

In addition:

- Nonrandomized but treated patients will be part of the safety population
- Randomized patients for whom it is unclear whether they took the IMP will be included in the safety population as randomized
- For patients on placebo but accidentally exposed to dupilumab, the treatment group allocation for as-treated analysis will be 300 q2w dose group

2.3.3 Pharmacokinetics (PK) population

The PK population will consist of all patients in the safety population with at least one evaluable functional dupilumab concentration result. Patients will be analyzed according to the treatment actually received.

2.3.4 Anti-drug antibody population

The anti-drug antibody (ADA) population will consist of all patients in the safety population with at least one reportable ADA results (either 'ADA negative' or 'ADA positive') after first dose of the study treatment. Patients will be analyzed according to the treatment actually received.

2.4 STATISTICAL METHODS

2.4.1 Demographics and baseline characteristics

Continuous data will be summarized using the number of available data, mean, standard deviation (SD), median, minimum, and maximum for each treatment group. Categorical and ordinal data will be summarized using the number and percentage of patients in each treatment group.

Parameters will be summarized on the randomized population analyzed in the treatment group to which they were randomized. Analyses for the safety population will be included in the appendices if the size of the safety population is different (>10%) from the size of that in the primary analysis population for any treatment group.

Parameters described in Section 2.1.1 will be summarized by treatment group and overall treatment groups using descriptive statistics.

Medical and surgical history will be summarized by treatment group and by system organ class (SOC) and preferred term (PT) sorted by internationally agreed order of SOC and by the

decreasing frequency of PT within SOC based on the overall incidence across treatment groups. Atopic medical history will be summarized separately.

No statistical testing on demographic and baseline characteristic data will be performed.

No specific description of the safety parameters will be provided at baseline. If relevant, the baseline values will be described along with each safety analysis.

2.4.2 Prior or concomitant medications

The prior and concomitant medications will be presented for the randomized population.

Medications will be summarized by treatment group according to the WHO-DD dictionary, considering the first digit of the anatomic category (ATC) class (anatomic category) and the first 3 digits of the ATC class (therapeutic category). All ATC codes corresponding to a medication will be summarized, and patients will be counted once in each ATC category (anatomic or therapeutic) linked to the medication. Therefore patients may be counted several times for the same medication.

The table for prior medications will be sorted by decreasing frequency of ATC followed by all other therapeutic classes based on the overall incidence across treatment groups. In case of equal frequency regarding ATCs (anatomic or therapeutic categories), alphabetical order will be used.

Concomitant medication received during first IMP to last IMP +14 days and concomitant medication received during first IMP to last IMP +98 days will be summarized separately. The tables for concomitant medications will be sorted by decreasing frequency of ATC followed by all other therapeutic classes based on the incidence in the dupilumab 300 q2w group. In case of equal frequency regarding ATCs (anatomic or therapeutic categories), alphabetical order will be used.

Medications will also be summarized by generic name sorted by decreasing frequency based on the overall incidence across treatment groups.

In addition, MFNS controller medication will be summarized separately. Asthma controller medication(s) will also be summarized for asthma patients at baseline, Week 24 and Week 48 visits.

2.4.2.1 INCS/MFNS controller medication

INCS/MFNS controller medication taken within 1 year prior to screening will be summarized by treatment group sorted by decreasing frequency of generic name followed by total daily dosage based on the incidence in the overall group.

The prescription of INCS/MFNS controller medication at randomization will be summarized regarding QD and BID by treatment group.

2.4.2.1.1 Compliance

During the study, the daily intake of the INCS/MFNS controller medication will be recorded on the electronic diary every morning by patients as number of actuations. Compliance for INCS will be calculated for each patient. For each day, a patient is considered as compliant to the prescribed controller medication if the actual dose is the same as or greater than the prescribed dose (ie, two actuations) for each nostril during each administration throughout the day: for one follows a QD (or BID) regimen, the patient is only considered as compliant to the prescribed INCS medication if the actual dose for each nostril administered during the day (or both in the morning and evening) is the same as or greater than the prescribed dose.

Percentage of Compliance is defined as the number of days when the patient is compliant to the prescribed controller medication divided by the number of days the patient stays in the treatment period (from first IMP to last IMP + 14 days).

2.4.3 Extent of investigational medicinal product exposure and compliance

The extent of IMP exposure and compliance will be assessed and summarized by actual treatment within the safety population (Section 2.3.2).

2.4.3.1 Extent of investigational medicinal product exposure

The extent of IMP exposure will be assessed by the duration of IMP exposure.

Duration of IMP exposure is defined as last dose date – first dose date + 14 days, regardless of unplanned intermittent discontinuations (see Section 2.5.3 for calculation in case of missing or incomplete data). Duration of IMP exposure will be summarized descriptively as a quantitative variable (number, mean, SD, median, minimum, and maximum). In addition, duration of treatment exposure will also be summarized categorically by numbers and percentages for each of the following categories and cumulatively according to these categories:

- >0 and ≤2 weeks
- >2 and ≤4 weeks
- >4 and ≤8 weeks
- >8 and ≤12 weeks
- >12 and ≤16 weeks
- >16 and ≤20 weeks
- >20 and ≤24 weeks
- >24 weeks

Additionally, the cumulative duration of treatment exposure will be provided, defined as the sum of the duration of treatment exposure for all patients, and will be expressed in patient years.

2.4.3.2 Compliance

A given administration will be considered noncompliant if the patient did not take the planned dose of treatment as required by the protocol. No imputation will be made for patients with missing or incomplete data.

Percentage of compliance for a patient will be defined as the number of administrations that the patient was compliant divided by the total number of administrations that the patient was planned to take during the treatment epoch defined in Section 2.1.4.

Treatment compliance percentages will be summarized descriptively as quantitative variables (number, mean, SD, median, minimum, and maximum). The percentage of patients whose compliance is <80% will be summarized in Section 2.4.4.5.

Cases of overdose (defined as at least twice the intended dose during an interval of less than 11 days) will constitute serious adverse events and will be listed as such. More generally, dosing irregularities will be listed in Section 2.2.1.

2.4.4 Analyses of efficacy endpoints

The primary analysis population for the efficacy endpoints will be the randomized ITT population which includes all patients who have been allocated to a randomized treatment regardless of whether the treatment kit was used or not. The efficacy analyses will be conducted according to the treatment to which they were randomized.

2.4.4.1 Analysis of primary efficacy endpoint(s)

Co-primary efficacy variables

The co-primary efficacy variables are: change from baseline in NC and in NPS (change from baseline in NC, NPSand LMK for Japan) at Week 24 assessed for dupilumab 300 mg q2w (Arm A) versus placebo (Arm B), respectively.

The following null hypothesis and alternative will be tested for dupilumab arm against placebo:

- H0: No treatment difference between the dupilumab dose regimen and placebo.
- H1: There is a treatment difference between the dupilumab dose regimen and placebo.

Primary statistical model (ITT analysis)

Each of the 2 co-primary efficacy endpoints (3 co-primary efficacy endpoints for Japan) will be analyzed using a hybrid method of the worst-observation carried forward (WOCF) and multiple imputation. Data collected after treatment discontinuation will be included in the analysis. With this approach, for patients who undergo surgery for NP or receive SCS for any reason, data collected post-surgery (actual date) or post SCS will be set to missing, and the worst post-baseline value on or before the time of surgery or SCS will be used to impute missing Week 24 value (for

patients whose post-baseline values are all missing, the baseline will be used to impute). For patients who discontinue the treatment without being rescued by surgery or receiving SCS, a multiple imputation approach will be used to impute missing Week 24 value, and this multiple imputation will use all patients who have not been rescued by surgery or receiving SCS at Week 24. Each of the imputed complete data will be analyzed by fitting an ANCOVA model with the baseline value of the corresponding co-primary endpoint, treatment group, asthma/NERD status, prior surgery history, and regions as covariates. Statistical inference obtained from all imputed data will be combined using Rubin's rule. Descriptive statistics including number of patients, mean, standard error, and least squares (LS) means will be provided. In addition, difference in LS means and the corresponding 95% confidence intervals (CI) will be provided along with the p-values.

Imputation and analysis model will be built with the following sample SAS code:

1. 40 datasets with a monotone missing pattern will be obtained, induced by Markov Chain Monte Carlo (MCMC) method on patients who have not been rescued by surgery or receiving SCS at Week 24.

```
proc mi data=dat_etd seed=14280 nimpute=40 out=dat_mc;
    mcmc impute=monotone;
    var asthma surgery region treatment basenps chg4nps chg8nps
    chg16nps   chg24nps;
run;
```

2. For each of the imputed dataset with monotone missing pattern in step 1, the remaining missing data will be imputed using the regression method for the monotone pattern with adjustment for covariates including treatment groups, region, asthma/NERD strata, surgery strata and baseline value of the response variable.

```
proc mi data=dat_mc nimpute=1 seed=14280 out=dat_mi;
    by _imputation_;
    class asthma surgery region treatment;
    monotone method=reg;
    var asthma surgery region treatment basenps chg4nps chg8nps chg16
    nps chg24nps;
run;
```

3. Each of the 40 imputed datasets will be merged with the one dataset imputed by WOCF approach, and then be analyzed using the main statistical model. These 40 imputed datasets will be saved.

```
%macro w1;
      %do i=1%to 40;
      data wocf&i.;
      set wocf;
      _imputation_=&i.;
```

```
      run;
      %end;
      data wocf_all;
      set %do j=1 %to 40; wocf&j. %end;;
      run;
%mend w1;

%w1

data dat_imp;
      set dat_mi  wocf_all;
Run;

proc sort data=dat_imp;
      by _imputation_;
run;

proc glm data= dat_imp;
      by _imputation_;
      class region asthma surgery treatment;
      model  = chg24nps = basenps asthma surgery region treatment;
      lsmeans treatment / stderr;
      estimate 'Diff Dupilumab vs Placebo' treatment -1 1;
      ods output LSMeans=implsmeans Estimates=implsmeandiff;
run;
```

4. Applying Rubin's rule to combine analysis results (point estimates and standard errors) from 40 imputations using PROC MIANALYZE for the LS means and difference in LS means between dupilumab and placebo. Sample code:

```
proc sort data=implsMeans; by trt01pn _imputation_;run;

proc mianalyze data= implsmeans;
      by trt01pn;
      modeleffects lsmean;
      stderr stderr;
      ods output ParameterEstimates=lsmeans;
run;

proc mianalyze data=implsmeandiff;
      modeleffects estimate;
      stderr stderr;
      ods output ParameterEstimates=lsmeandiff;
run;
```

Sensitivity analysis

For all sensitivity approaches, for patients who undergo surgery for NP or receive SCS for any reason, data collected post-surgery or post SCS will be set to missing.

Mixed-effect model with repeated measures (MMRM) approach

The model will include change from baseline values up to week 24 as response variables, and factors (fixed effects) for treatment, stratification factor (co-morbid asthma/NERD, prior surgery, region), visit, treatment-by-visit interaction, NPS/NC baseline value and baseline-by-visit interaction. An unstructured correlation matrix will be used to model the within-patient errors. Parameters will be estimated using restricted maximum likelihood method with the Newton-Raphson algorithm. The DDFM=KR will be used to compute the denominator degrees of freedom in the F-test. Data collected after treatment discontinuation will be included in the analysis. Descriptive statistics including number of patients, mean, standard error and LS means will be provided. In addition, difference in LS means, the corresponding 95% CI and the p-value will be provided for comparison of dupilumab against placebo. No imputation will be performed for the MMRM model.

If MMRM model fails to achieve convergence due to complexity of model specification, we will add some procedures to handle this issue including: 1) using maximum likelihood estimation instead of restricted maximum likelihood method; 2) covariance structure selection to reduce the number of unknown parameters to be estimated, the path is UN→TOEPH→TOEP→AR(1).

Pattern mixture model with copy increment from placebo

Each of the 2 co-primary efficacy endpoints (3 co-primary efficacy endpoints for Japan) will be analyzed with imputed missing value at 24 weeks using pattern mixture model with copy increment from placebo (12). This copy increment from placebo implies that when subjects discontinue treatment early, they continue to take advantage of their previous therapy, but they progress in the same way as subjects in the placebo group.

The imputed dataset will be analyzed by fitting an ANCOVA model same as the one in primary analysis. Descriptive statistics including number of patients, mean, standard error, and least squares (LS) means will be provided. In addition, difference in LS means and the corresponding 95% confidence intervals (CI) will be provided along with the p-values.

Tipping point analysis

Each of the 2 co-primary efficacy endpoints (3 co-primary efficacy endpoints for Japan) will be analyzed with imputed missing value at 24 weeks as follows.

- **Step 1.** Monotone missing pattern was induced by Markov Chain Monte Carlo (MCMC) method using PROC MI: for patients who have intermediate missing values, the intermediate missing values will be imputed assuming a multivariate normal distribution over observations from all visits. 40 datasets with a monotone missing pattern will be obtained using this method.

- **Step 2.** For each of the imputed dataset with monotone missing pattern obtained in Step 1, the remaining missing data will be imputed using the regression method for the monotone pattern with adjustment for covariates including response variable, treatment groups, asthma/NERD strata, prior surgery strata, region, and baseline value of the corresponding endpoint. All available data in the monotone missing pattern data will be used. One imputed dataset will be obtained for each of the imputed dataset at Step 1. So, 40 fully imputed datasets will be obtained altogether.
- **Step 3.** The imputed values in Dupilumab group are added by a positive amount *d* for each imputed data sets.
- **Step 4.** The imputed values in placebo group are subtracted by a positive amount *p* for each imputed data sets.
- **Step 5.** Change from baseline in endpoint will be analyzed using ANCOVA model same as the one in primary analysis. Then the SAS MIANALYZE procedure will be used to generate statistical inferences by combining results from the 40 analyses using Rubin's formula.

Step 3 to Step 5 will be repeated iteratively until the p-value for treatment effect of dupilumab compared to placebo estimated in Step 5 is >0.05.

LS mean difference between dupilumab and placebo in change from baseline in co-primary endpoints at Week 24 and the corresponding p-values will be provided for each combination of shift parameters.

Additional analysis

An additional analysis will be conducted on the 2 co-primary efficacy endpoints (3 co-primary efficacy endpoints for Japan) which includes all data including data collected after SCS for any reason and treatment discontinuation but excludes post NP surgery data. For patients who undergo surgery for NP, data collected post-surgery will be set to missing (as a comparison, data collected post-surgery and post SCS for any reason are both set to missing in the primary approach), and the worst post-baseline value on or before the time of surgery will be used to impute missing Week 24 value (for patients whose post-baseline values are all missing, the baseline will be used to impute). For patients who discontinue the treatment without being rescued by surgery, all data collected after treatment discontinuation or after SCS for any reason will be included in the analysis and a multiple imputation approach will be used to impute missing Week 24 value, and this multiple imputation will use all patients who have not been rescued by NP surgery at Week 24. The data will be analyzed in the same ANCOVA model for the primary approach. Descriptive statistics including number of patients, mean, standard error, and least squares (LS) means will be provided. In addition, difference in LS means and the corresponding 95% confidence intervals (CI) will be provided along with the p-values.

Mixed-effect model with repeated measures (MMRM) approach for NC as binary response data

In the primary analysis, NC will be analyzed as the average of 28-day nasal congestion data. To assess the robustness of this approach with respect to the ordinal categorical data nature of nasal

congestion score, a MMRM approach on NC as longitudinal binary response data which gives statistical inference in terms of rate difference will be performed based on methods proposed and evaluated by Fan (13). First, the data will be processed to a repeated measures binary data format to facilitate the use of methods for longitudinal binary response data as follows.

A) Baseline: the baseline of the NC will be the median of the daily score in the 7 days prior to randomization if there are at least 4 observations during these 7 days; if there are less than 4 observations, the baseline of the NC will be the median of the last 4 daily scores prior to randomization. In both cases, if the median ends in a .5 score, round it up.

B) Justification: The primary analysis uses the mean. Assuming the NC symptoms are reasonably stable during these 7 days (observed in ACT12340), it deemed to be appropriate to use median here. Using median instead of mean is also in accordance with the ordinal categorical data nature of nasal congestion score.

C) Week 24: for each day during study day 142 to study day 169, the change from baseline in scores will be re-coded as

   a) =1 if a subject improves at least 1 category at that day.

   b) =0 if a subject does not improve at least 1 category at that day.

For a patient with observations during days 142 to 169 (28 days), he would have daily symptom improvement score as a 28-dimensional vector. Each element of the vector represents whether the subject shows improved symptom in nasal congestion on that day compared with baseline. Missing data in any day during study day 142 to study day 169 will be imputed as not improved for that day (=0).

By such an approach, daily data are used in the model, and such data also preserve the categorical data nature of nasal congestion score. This approach gives statistical inference in terms of rate difference and has been theoretically and quantitatively validated (13).

In this model, factors (fixed effects) for treatment, stratification factors (region, co-morbid asthma/NERD and prior surgery), day, treatment-by-day interaction, baseline value as a categorical variable will be included. Generalized estimating equations (GEE) will be used to obtain statistical inference. A variance component working covariance matrix which is merely a weight matrix will be used in the model to obtain unweighted proportions. Descriptive statistics including number of patients, LS average proportion of patients with improvement across the 28 days, and corresponding standard error will be provided. In addition, difference in LS average proportion of patients with improvement, the corresponding 95% CI and the p-value will be provided for comparison of dupilumab against placebo.

Subgroup analyses

To assess the consistency in treatment effects across different subgroup levels, subgroup analyses will be conducted for the co-primary efficacy endpoints with respect to

- Age group (18-64, ≥65 years)
- Gender (Male, Female)

- Baseline weight (<70, ≥70- < 90, ≥ 90 kg; <60, ≥60 kg)
- Baseline BMI (<25, ≥25-<30, ≥30 kg/m^2)
- Region (East Europe: Russia, Ukraine, Cz. Republic, Hungary, Poland, Romania, Bulgaria; Western Countries: France, Germany, Italy, United Kingdom, Netherlands, USA)
- Territory (North America: USA; European Union: Cz. Republic, Hungary, Poland, Romania, Bulgaria, France, Germany, Italy, United Kingdom and Netherlands; Rest of World: Russia and Ukraine)
- Prior NP surgery history (Yes, No)
- Race (Caucasian/White, Black/of African descent, Asian/Oriental, all the Others)
- Ethnicity (Hispanic, non-Hispanic)
- Asthma comorbidity (Yes, No)
- Asthma and/or NERD (Yes, No)
- NERD (Yes, No)
- Allergic rhinitis at baseline (Yes, No)
- SCS use during the past 2 years prior to V1 (Yes, No)

To assess the consistency of the treatment effects across the subgroup levels, subgroup analyses will be conducted for co-primary endpoints at Week 24. The analysis will be performed based on imputed datasets from the primary analysis.

To test the interaction between treatment and subgroup factor, an ANCOVA model incorporating subgroup-by-treatment interaction will be built for each subgroup factor. The model will include all the covariates in the main statistical model plus the subgroup variable (if not one of the covariates adjusted in the main model already) and the subgroup-by-treatment interaction. Statistical inference obtained from all imputed data will be combined using Rubin's rule. A p-value for the test of interaction will be provided based on the combined inference.

In each subgroup, the co-primary endpoints will be analyzed using the primary approach for co-primary endpoints, but on the specific subgroup of the imputed primary analysis population. Descriptive statistics including number of patients, mean, standard error, and least squares (LS) means for each subgroup will be provided. In addition, difference in LS means and the corresponding 95% confidence intervals (CI) will be provided for each subgroup.

2.4.4.2 Analyses of key secondary efficacy endpoints

2.4.4.2.1 Analysis of the change from baseline in LMK, TSS, UPSIT score, the daily loss of smell assessment, and SNOT-22 scores at Week 24 for dupilumab 300 mg q2w versus placebo

The change from baseline in LMK score, TSS, UPSIT score, daily loss of smell, and SNOT-22 at Week 24 will be assessed for dupilumab 300 mg q2w (Arm A) versus placebo (Arm B) and will be analyzed using the hybrid method of the WOCF and the MI the same as the primary approach of the co-primary endpoints. Descriptive statistics including number of patients, mean, standard error, and least squares (LS) means will be provided. In addition, difference in LS means and the corresponding 95% confidence intervals (CI) will be provided along with the p-values.

Note: LMK will not be a secondary endpoint for Japan as it is already a co-primary endpoint.

2.4.4.2.2 Analysis of proportion and time-to-event of patients requiring rescue treatment defined as: use systemic corticosteroids or NP surgery (actual or planned) during the treatment period for dupilumab 300 mg q2w versus placebo

Proportion of patients requiring rescue treatment defined as: use systemic corticosteroids or NP surgery (actual or planned) during the treatment period will be derived and analyzed using the Cox proportional hazards model. The event of interest is the first SCS use for any reason or NP surgery (actual or planned), whichever is earlier if both occur. The decision date of surgery for nasal polyps or the first SCS intake date will be used as the event date, whichever is earlier if both occur. In case the decision date of NP surgery is missing, the planned surgery date will be used in place. In case the planned surgery date is also missing, the actual surgery date will be used in place.

Due to potential low number of events in each single study, proportion of patients requiring rescue treatment defined as: use systemic corticosteroids or NP surgery (actual or planned) during the treatment period for dupilumab 300 mg q2w versus placebo will be analyzed by pooling the two NP studies SINUS-52 (EFC14280) and SINUS-24 (EFC14146, the study this SAP is for). In the pooled analysis, the treatment period by dupilumab 300 mg q2w will be the total of the entire treatment period in Arm A of EFC14280, the treatment period of 300 mg q2w in Arm B (that is, the first 24 weeks) of EFC14280, and the entire treatment period in Arm A of EFC14146; and the treatment period by placebo will be the total of the entire treatment period in Arm C of EFC14280 and the entire treatment period in Arm B of EFC14146.

The pooled data will be analyzed by the Cox model. The Cox model will include the event as the dependent variable, and a study indicator (0 for EFC14280 and 1 for EFC14146), treatment group, asthma/NERD strata, prior surgery strata, and region (pooled countries) as covariates. Hazard ratio and corresponding 95% CI and p values will be estimated for dupilumab 300 mg q2w versus placebo. The Kaplan-Meier method will be used to derive the probabilities that a patient would experience events up to Week 52 for 300 mg q2w and placebo. Kaplan-Meier curves will be generated; quartiles and point probabilities will be calculated. Interval estimates will be calculated using 95% point wise confidence intervals.

2.4.4.2.3 Analysis of the change from baseline in FEV1 at Week 24 for asthma patients

The change from baseline in FEV1(L) at Week 24 will be assessed by pooling the two NP studies SINUS-52 (EFC14280) and SINUS-24 (EFC14146, the study this SAP is for). The missing data in FEV1 at Week 24 will be imputed using the hybrid method of the WOCF and the MI the same as the primary approach of the co-primary endpoints. Each of the imputed complete data will be analyzed by fitting an ANCOVA model with the change from baseline in FEV1 at Week 24 as the response variable, and a study indicator (0 for EFC14280 and 1 for EFC14146), the baseline value of FEV1, treatment group, prior surgery strata, and region (pooled countries) as covariates. Statistical inference obtained from all imputed data will be combined using Rubin's rule. Descriptive statistics including number of patients, mean, standard error, and least squares (LS) means will be provided. In addition, difference in LS means and the corresponding 95% confidence intervals (CI) will be provided along with the p-values.

2.4.4.3 Multiplicity issues

The multiplicity procedure is proposed to control the overall type-I error rate for testing the co-primary and selected secondary endpoints. The overall alpha is 0.05. The comparisons with placebo will be tested based on the hierarchical order below at 2-sided $\alpha = 0.05$:

1. Co-primary efficacy endpoints:

In countries other than Japan:

- Change from baseline in NC and in NPS at Week 24 for 300mg q2w versus placebo.

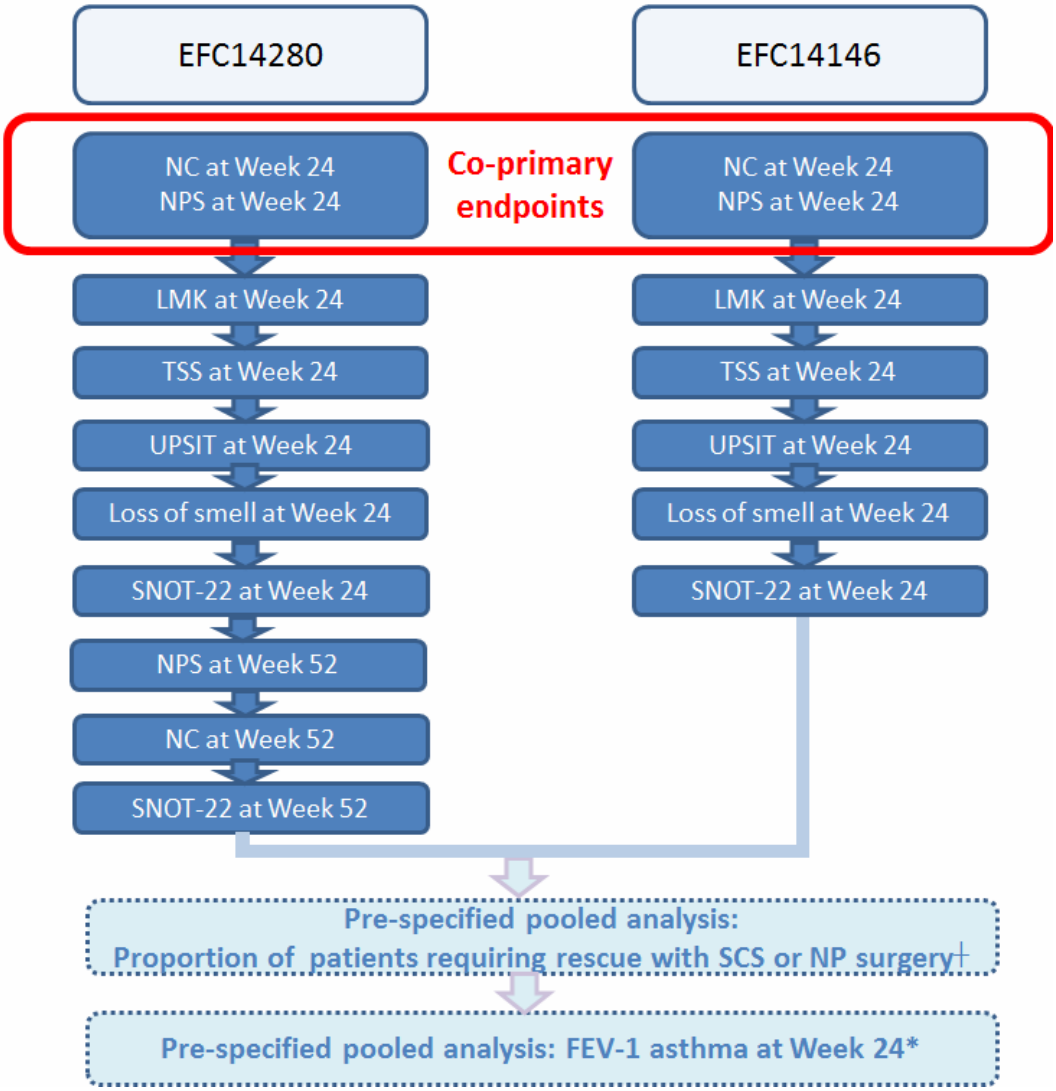
In Japan:

- Change from baseline in NC, in NPS, and in CT LMK at Week 24 for 300mg q2w versus placebo.

2. Secondary efficacy endpoints:

- Change from baseline in CT LMK score at Week 24 for 300mg q2w versus placebo (this will not be a secondary endpoint for Japan as it is already a co-primary endpoint).
- Change from baseline in TSS at Week 24 for 300mg q2w versus placebo.
- Change from baseline in UPSIT at Week 24 for 300mg q2w versus placebo.
- Change from baseline in loss of smell daily symptoms at Week 24 for 300mg q2w versus placebo.
- Change from baseline in SNOT-22 at Week 24 for 300mg q2w versus placebo.
- Proportion of patients requiring rescue treatment defined as: use SCS or NP surgery (actual or planned) during the treatment period for 300mg q2w versus placebo using the pooled data from the EFC14280 and EFC14146.†
- Change from baseline in FEV1(L) at Week 24 for 300mg q2w versus placebo for patients with co-morbid asthma using the pooled data from the EFC14280 and EFC14146.*

The figure below shows individual studies hierarchy for EFC14280 and EFC14146 as well as the pre-specified pooled analysis for proportion of patients requiring rescue treatment with SCS or NP surgery and change from baseline in FEV1(L) at Week 24.



┼: The pooled analysis for this endpoint (Proportion of patients requiring rescue treatment of SCS or NP surgery) will be tested in the hierarchy only when in both EFC14280 and EFC14146 hierarchies, all endpoints before this one reach statistical significance with p-value ≤ 0.05.

*: The pooled analysis for this endpoint (Change from baseline in FEV1 at Week 24) will be tested in the hierarchy only when the previous endpoint (the pooled analysis for the proportion of patients requiring rescue treatment of SCS or NP surgery) achieves statistical significance in this hierarchical testing procedure as described in the previous paragraph.

In countries other than Japan, the study is considered positive when both co-primary endpoints, the change from baseline in NC and in NPS at Week 24, achieve statistical significance.

In Japan, the study is considered positive when all co-primary endpoints, the change from baseline in NC, in NPS, and in CT LMK at Week 24, achieve statistical significance.

2.4.4.4 Additional efficacy analyses

2.4.4.4.1 Comparisons for other secondary efficacy endpoints

- Comparisons at Week 24 will be made between dupilumab 300 mg q2w (Arm A) versus placebo (Arm B).
- Comparisons at Week 48 will be made between dupilumab 300 mg q2w (Arm A) versus placebo (Arm B).
- Comparisons will be made for the following secondary endpoints:
  - Change from baseline and time course profiles in NPS, NC, LMK, TSS, UPSIT, daily assessed loss of smell, and SNOT-22 at Week 48,
  - Change from baseline at Week 24 in:
  - VAS for overall rhinosinusitis,
  - NPIF,
  - VAS for EQ-5D
  - In the severity of rhinorrhea (anterior/posterior nasal discharge) daily symptom score assessed by the patient,
  - Percentage change from baseline at Week 24 and time course profiles in:
  - NC
  - Daily assessed loss of smell
  - TSS
  - Percentage change from baseline at Week 24 in VAS for overall rhinosinusitis
  - Proportion of responders at Week 24 (defined as patients with improvement by at least 1 point in NPS),
  - Proportion of responders at Week 24 (defined as patients with improvement by at least 2 points in NPS),
  - Proportion of patients with improvement by at least 1 point in NPS and 0.5 reductions in NC at Week 24 and Week 48,
  - Proportion of patients with greater than or equal to the minimal clinically important difference (MCID)(≥8.9) in SNOT-22 at Week 24,
  - Proportion of patients with overall rhinosinusitis severity VAS ≤7 at Week 24.
  - The cumulative distribution function (CDF) of change from baseline in NC, NPS, SNOT-22 and VAS for rhinosinusitis severity at Week 24,
  - Proportion of patients with anosmia by UPSIT scores at Week 24.

2.4.4.4.2 Analyses of other efficacy endpoints for 300 mg q2w versus placebo at Week 24

The change from baseline in VAS for overall rhinosinusitis, NPIF, EQ-5D utility and daily assessed severity of rhinorrhea at Week 24 and percent change from baseline in NC, TSS, daily assessed loss of smell and VAS for overall rhinosinusitis at Week 24 will be assessed for dupilumab 300 mg q2w (Arm A) versus placebo (Arm B) and will be analyzed using the hybrid method of the WOCF and the MI the same as the primary approach of the co-primary endpoints. Descriptive statistics including number of patients, mean, standard error, and least squares (LS) means will be provided. In addition, difference in LS means and the corresponding 95% confidence intervals (CI) will be provided along with the p-values.

For responder type endpoints including

- The proportion of responders at Week 24 (defined as patients with improvement by at least 1 point in NPS)
- The proportion of responders at Week 24 (defined as patients with improvement by at least 2 points in NPS)
- The proportion of patients with improvement by at least 1 point in NPS and 0.5 reductions in NC at Week 24
- The proportion of patients with MCID $\geq$8.9 in SNOT-22 at Week 24
- The proportion of patients with VAS $\leq$7 at Week 24

The Cochran-Mantel-Haenszel test will be performed on the association between the responder status and treatment group, stratified by asthma/NERD status, prior surgery history, and region. Comparisons of the proportions of responders between dupilumab 300mg q2w and placebo will be derived. And corresponding odds ratios and 95% CI will be reported. Patients who are indicated for surgery for NP or receive SCS for any reason will be considered as non-responders for time points after using SCS or surgery. For patients who discontinue treatment without using SCS or surgery, data collected during the off- treatment period will be used to determine the responder/non- responder status. Missing data at Week 24 will be considered as non-responders.

In addition, the cumulative distribution function (CDF) of change from baseline in NC, NPS, SNOT-22 and VAS for rhinosinusitis severity at Week 24 will also be presented by treatment.

2.4.4.4.3 Analyses of efficacy endpoints for Arm A versus Arm B at Week 48

The change from baseline in NPS, NC, LMK, TSS, UPSIT, daily assessed loss of smell, SNOT-22, VAS for overall rhinosinusitis, and daily assessed severity of rhinorrhea at Week 48 and percent change from baseline in NC, TSS, daily assessed loss of smell and VAS for overall rhinosinusitis at Week 48 will be assessed for Arm A versus Arm B and only descriptive statistics will be provided by treatment. For analyses at week 48, dupilumab 300 mg q2w (Arm A) versus placebo (Arm B) will be analyzed using the hybrid method of the WOCF and the MI in the same fashion as for the primary analysis for co-primary endpoints at Week 24. Descriptive statistics including number of patients, mean, standard error, and least squares (LS) means will be provided.

In addition, difference in LS means and the corresponding 95% confidence intervals (CI) will be provided.

For responder type endpoints including

- The proportion of patients with improvement by at least 1 point in NPS and 0.5 reductions in NC at Week 48

The Mantel-Haenszel estimator will be provided on the association between the responder status and treatment group, stratified by asthma/NERD status, prior surgery history, and region. Comparisons of the proportions of responders between dupilumab 300mg q2w (Arm A) and placebo (Arm B) will be derived. And corresponding odds ratios and 95% CI will be reported. Patients who are indicated for surgery for NP or receive SCS for any reason will be considered as non-responders for time points after using SCS or surgery. For patients who discontinue treatment without using OCS or surgery, data collected during the off- treatment period will be used to determine the responder/non- responder status. Missing data at Week 48 will be considered as non-responders.

2.4.4.4.4 Time course profile of the efficacy endpoints over 24 and 48 weeks

For the continuous efficacy endpoints, descriptive statistics including number of patients, mean, standard error, and least squares (LS) means will be provided and plotted separately by treatment group for 300mg q2w (Arm A) versus placebo (Arm B) at each visit up to Week 48. In addition, using the hybrid method of the worst-observation carried forward (WOCF) and the multiple imputation separately at each visit, difference in LS means and the corresponding 95% confidence intervals (CI) will be provided along with the p-values (p-values will be provided up to Week 24). Such analyses will be conducted on the following variables:

- Change from baseline in NPS
- Change and percent change from baseline in NC
- Change from baseline in CT LMK scores
- Change and percent change from baseline in TSS
- Change from baseline in UPSIT
- Change and percent change from baseline in loss of smell
- Change from baseline in SNOT-22

For proportion type efficacy endpoints, the response rate at each visit up to Week 48 will be summarized and plotted by treatment groups. The odds ratio and the corresponding 95% CI will be derived between dupilumab 300mg q2w and placebo at each visit with p-values provided up to Week 24. Patients who are indicated for surgery for NP or receive SCS for any reason will be considered as non-responders for time points after using SCS or surgery. For patients who discontinue treatment without using SCS or surgery, data collected during the off- treatment period will be used to determine the responder/non- responder status. Missing data will be considered as non-responders. Such analyses will be conducted on the following variables:

- Proportion of responders (defined as patients with improvement by at least 1 point in NPS),
- Proportion of responders (defined as patients with improvement by at least 2 point in NPS),
- Proportion of patients with improvement by at least 1 point in NPS and 0.5 reductions in NC
- Proportion of patients with MCID≥8.9 in SNOT-22
- Proportion of patients with VAS ≤7

2.4.4.4.5 Analysis of anosmia category

For UPSIT, number (%) of patients with anosmia will be summarized for dupilumab 300mg q2w and placebo, respectively, at baseline and Week 24.

The anosmia category is defined as:

- 0-18 anosmia
- 19-25 severe microsmia
- 26-30 moderate microsmia
- 31-34 mild microsmia
- 35-40 normal

2.4.4.5 Analyses of exploratory efficacy endpoints

Total SCS dose prescribed (in mg) during the study period and total SCS intake in days and courses during the study period will be summarized by treatment groups. A course of SCS is considered continuous if treatment is separated by less than 7 days. Various doses of systemic corticosteroids will be converted to prednisone-equivalent oral corticosteroid. For patients who discontinue treatment, data collected during the off- treatment period will also be used. Descriptive statistics including number of subjects, mean, standard error, min, and max will be provided, and no statistical testing will be conducted.

Proportion and time-to-event of patients requiring rescue treatment with SCS for any season, SCS for nasal polyps, SCS for respiratory disease, and surgery (actual or planned) for NP will also be performed using the same method as those for the secondary endpoint of the proportion of patients requiring rescue with SCS or NP surgery (actual or planned) during the treatment period for dupilumab 300 mg q2w versus placebo using the pooed data of the two NP studies SINUS-52 (EFC14280) and SINUS-24 (EFC14146).

Change from baseline in decreased/loss of sense of taste symptom severity and SNOT-22 items: “decreased sense of smell/taste”, “difficulty falling asleep”, ”wake up at night”, “lack of a good night's sleep”, “wake up tried”, ”fatigue”, and “reduced productivity”, (FEV1, FVC and FEF25-75 in patients with asthma) at Week 24 and Week 48 will be assessed separately. For analyses at

Week 24, Dupilumab 300 mg q2w (Arm A) versus placebo (Arm B) will be analyzed using the hybrid method of the WOCF and the MI in the same fashion as for the primary analysis for co-primary endpoints at Week 24. For analyses at Week 48, Arm A versus Arm B will be analyzed using descriptive statistics. Descriptive statistics including number of patients, mean, standard error, and least squares (LS) means will be provided. In addition, difference in LS means and the corresponding 95% confidence intervals (CI) will be provided.

2.4.4.6 Missing data handling

For all continuous efficacy endpoints, the details of missing data handling are described in Section 2.4.4.1.

For responder type endpoints, the details of missing data handling are described in Section 2.4.4.4.2.

In addition, the reason and pattern of missing data will be carefully examined and pattern mixture modeling approach and other sensitivity analyses will also be performed as specified in Section 2.4.4.1.

2.4.5 Analyses of safety data

The summary of safety results will be presented by treatment group.

General common rules

All safety analyses will be performed on the safety population as defined in Section 2.3.2, unless otherwise specified, using the following common rules:

- The baseline value is defined as the last available value before the first dose of IMP
- The potentially clinically significant abnormality (PCSA) values are defined as abnormal values considered medically important by the Sponsor according to predefined criteria/thresholds based on literature review and defined by the Sponsor for clinical laboratory tests, vital signs, and ECG [Appendix A])
- PCSA criteria will determine which patients had at least 1 PCSA during the treatment-emergent adverse event period, taking into account all evaluations performed during the treatment-emergent adverse event period, including nonscheduled or repeated evaluations. The number of all such patients will be the numerator for the on-treatment PCSA percentage
- The treatment-emergent PCSA denominator by group for a given parameter will be based on the number of patients assessed for that given parameter in the treatment-emergent adverse event period by treatment group on the safety population
- For quantitative safety parameters based on central laboratory/reading measurements, descriptive statistics will be used to summarize results and change from baseline values by visit and treatment group. Summaries will include the endpoint value and/or the worst on-treatment value. The endpoint value is commonly defined as the value collected at the end

of treatment. If this value is missing, this endpoint value will be the closest value prior to the end of treatment epoch. The worst value is defined as the nadir and /or the peak post-baseline (up to the end of treatment epoch or EOT) according to the direction (minimum or maximum) of the abnormality as defined in the PCSA list

- The analysis of the safety variables will be essentially descriptive and no systematic testing is planned. Relative risks versus placebo and their 95% confidence intervals may be provided, if relevant
- All safety values including unscheduled measurements will be assigned to the appropriate safety analysis visit window defined in Section 2.5.3.

2.4.5.1 Analyses of adverse events

Generalities

The primary focus of adverse event reporting will be on treatment-emergent adverse events. Pretreatment and post-treatment adverse events will be described separately.

If an adverse event date/time of onset (occurrence, worsening, or becoming serious) is incomplete, an imputation algorithm will be used to classify the adverse event as pretreatment, treatment-emergent, or post-treatment. The algorithm for imputing date/time of onset will be conservative and will classify an adverse event as treatment emergent unless there is definitive information to determine it is pretreatment or post-treatment. Details on classification of adverse events with missing or partial onset dates are provided in Section 2.5.2.

Adverse event incidence tables will present by SOC, HLGT, HLT, and PT, sorted in alphabetical order for each treatment group, the number (n) and percentage (%) of patients experiencing an adverse event. Multiple occurrences of the same event in the same patient will be counted only once in the tables within a treatment phase. The denominator for computation of percentages is the safety population within each treatment group.

Sorting within tables ensures the same presentation for the set of all adverse events within the observation period (pretreatment, treatment-emergent, and post-treatment). For that purpose, the table of all treatment-emergent adverse events presented by SOC and PT sorted by the internationally agreed SOC order and decreasing frequency of PTs within SOCs will define the presentation order for all other tables unless otherwise specified. Sorting will be based on results for the dupilumab 300 mg q2w group.

Analysis of all treatment-emergent adverse events

The following treatment-emergent adverse event summaries will be generated for the safety population.

- Overview of treatment-emergent adverse events, summarizing number (%) of patients with any
  - Treatment-emergent adverse event

  - Serious treatment-emergent adverse event
  - Treatment-emergent adverse event leading to death
  - Treatment-emergent adverse event leading to permanent treatment discontinuation
- All treatment-emergent adverse event by primary SOC, HLGT, HLT, and PT, showing number (%) of patients with at least 1 treatment-emergent adverse event sorted by the SOC internationally agreed order. The other levels (HLGT, HLT, PT) will be presented in alphabetical order
- Number (%) of patients experiencing TEAE(s) presented by PT, sorted by decreasing incidence of PT
- All treatment-emergent adverse events by primary SOC and PT, showing the number (%) of patients with at least 1 treatment-emergent adverse event, sorted by the internationally agreed SOC order and by decreasing incidence of PTs within each SOC in the 300 q2w treatment group
- All treatment-emergent adverse events regardless of relationship and related by primary SOC, HLGT, HLT and PT, showing the number (%) of patients with at least 1 treatment-emergent adverse event, sorted by the internationally agreed SOC order. The other levels (HLGT, HLT, PT) will be presented in alphabetical order
- All treatment-emergent adverse events by maximal severity, presented by primary SOC and PT, showing the number (%) of patients with at least 1 treatment-emergent adverse event by severity (ie, mild, moderate, or severe), sorted by the internationally agreed SOC order and by decreasing incidence of PTs within each SOC in the 300 q2w treatment group
- Number (%) of patients experiencing treatment-emergent adverse event(s) presented by primary and secondary SOC, HLGT, HLT, and PT sorted by the internationally agreed SOC order. The other levels (HLGT, HLT, PT) will be presented in alphabetical order.

A listing of all treatment-emergent adverse events will be presented.

Analysis of all treatment emergent serious adverse event(s)

- All treatment-emergent serious adverse events by primary SOC, HLGT, HLT, and PT, showing the number (%) of patients with at least 1 serious treatment-emergent adverse event, sorted by the internationally agreed SOC order. The other levels (HLGT, HLT, PT) will be presented in alphabetical order
- All treatment-emergent serious adverse events regardless of relationship and related to IMP, by primary SOC, HLGT, HLT, and PT, showing the number (%) of patients with at least 1 treatment-emergent serious adverse event, sorted by the internationally agreed SOC order. The other levels (HLGT, HLT, PT) will be presented in alphabetical order
- All treatment-emergent serious adverse events by primary SOC and PT, showing the number (%) of patients with at least 1 treatment-emergent serious adverse event, sorted by the internationally agreed SOC order and by decreasing incidence of PTs within each SOC in the 300 q2w treatment group.

A listing of all treatment-emergent serious adverse events will be presented.

Analysis of all treatment-emergent adverse event(s) leading to treatment discontinuation

- All treatment-emergent adverse events leading to treatment discontinuation, by primary SOC, HLGT, HLT, and PT, showing the number (%) of patients sorted by the internationally agreed SOC order. The other levels (HLGT, HLT, PT) will be presented in alphabetical order
- All treatment-emergent adverse events leading to permanent treatment discontinuation or drug interruption, by primary SOC and PT, showing number (%) of patients with at least one TEAE, will be presented by SOC internationally agreed order and by decreasing incidence of PTs within each SOC in the 300 q2w treatment group.

Analysis of adverse events of special interests (AESI) and other selected AE groupings

- All treatment-emergent adverse events, by selected standardized MedDRA query (SMQ) and PT or by laboratory values (as in ALT elevation), showing the number (%) of patients with at least 1 PT, sorted by decreasing incidence of PTs within each SMQ in the 300 q2w treatment group.
- For each AESI and other selected AE groupings,
  - Number (%) of patients with any specific TEAE
  - Number (%) of patients with any specific serious AE (regardless of treatment emergent status)
  - Number (%) of patients with any specific treatment emergent serious AE
  - Number (%) of patients with any specific AE leading to death
  - Number (%) of patients with any specific TEAE leading to permanent study drug discontinuation
  - Number (%) of patients with any specific TEAE by maximum intensity, corrective treatment, and final outcome
  - Number of any specific TEAE adjusted by the exposure duration
  - Number of patients with any specific TEAE adjusted by the exposure duration at risk. For each specific TEAE, Kaplan-Meier estimates of cumulative incidence at Week 12, 24, 36 and 48 and K-M plot may be provided to depict the course of onset over time if the number of events is large enough.
  - Number (%) of patients with injection site reactions by the related injection.
  - Number (%) of patients with different number of injection site reactions.
- In addition, AESIs reported by the investigator in CRF will be summarized separately.

Analysis of treatment-emergent adverse events up to Week 24 (On treatment adverse event)

- Selective safety analyses will be performed on safety data from baseline to Week 24 with respect to TEAE, SAE, TEAE leading to permanent treatment discontinuation or drug

interruption, AE adverse events of special interests (AESI) and other selected AE groupings.

Analysis of pretreatment adverse events

- All pretreatment adverse events by primary SOC and PT, showing the number (%) of patients with at least 1 pretreatment adverse event, sorted by the internationally agreed SOC order and decreasing incidence of PTs within each SOC in the 300 q2w treatment group.
- All serious pretreatment adverse events by primary SOC and PT, showing the number (%) of patients with at least 1 pretreatment adverse event, sorted by the internationally agreed SOC order and decreasing incidence of PTs within each SOC
- All pretreatment adverse events leading to permanent treatment discontinuation by primary SOC and PT, showing the number (%) of patients, sorted by the internationally agreed SOC order and decreasing incidence of PTs within each SOC
- All pretreatment adverse events leading to death by primary SOC and PT, showing the number (%) of patients, sorted by the internationally agreed SOC order and decreasing incidence of PTs within each SOC
- If only a few patients have pretreatment adverse events leading to permanent treatment discontinuation, a listing will be presented instead of the summary table above.

2.4.5.2 Deaths

The following summaries of deaths will be generated for the safety population.

- Number (%) of patients who died by study period (on-study, on-treatment, post-study)
- Deaths in nonrandomized patients or randomized but not treated patients
- Treatment-emergent adverse events leading to death (death as an outcome on the adverse event case report form page as reported by the Investigator) by primary SOC , HLGT, HLT, and PT showing number (%) of patients sorted by internationally agreed SOC order, with HLGT, HLT, and PT presented in alphabetical order within each SOC
- All pretreatment adverse events leading to death by primary SOC and PT, showing the number (%) of patients, sorted by the internationally agreed SOC order and decreasing incidence of PTs within each SOC.

A listing of deaths will be provided.

2.4.5.3 Analyses of laboratory variables

The summary statistics (including number, mean, median, Q1, Q3, standard deviation, minimum and maximum) of all laboratory variables (central laboratory values and changes from baseline) will be calculated for each visit or study assessment (baseline, each post-baseline time point, last on-treatment and/or worst on-treatment value) by treatment group. For each continuous parameters listed in Section 2.1.4.3, mean changes from baseline with the corresponding standard

error will be plotted over time in each treatment group. This section will be organized by biological function as specified in Section 2.1.4.3.

The incidence of PCSAs (list provided in Appendix A) at any time during the treatment-emergent adverse event period will be summarized by biological function and treatment group whatever the baseline level and/or according to the following baseline status categories:

- Normal/missing
- Abnormal according to PCSA criterion or criteria

The incidence of PCSAs from baseline to Week 24 during the treatment-emergent adverse event period will be summarized by biological function and treatment group whatever the baseline level and/or according to the following baseline status categories:

- Normal/missing
- Abnormal according to PCSA criterion or criteria

For parameters for which no PCSA criteria are defined, similar table(s) using the normal range will be provided.

For PCSA analyses, the laboratory measurements obtained at either scheduled or unscheduled visits should be used; and, both the centralized and local test results should be used, as long as their available dates/time is different from each other's. Centralized data will be used preferentially to the local measures in the PCSA analyses when measurements are performed on the same date and at the same time for a given laboratory test.

Possible Drug-induced liver injury

If there is any imbalance in the incidence of liver-related adverse events across the treatment groups, analysis of liver-related adverse events will be performed.

The liver function tests, namely AST, ALT, alkaline phosphatase, and total bilirubin, are used to assess possible drug-induced liver toxicity. The proportion of patients with PCSA values at any post-baseline visit by baseline status will be displayed by treatment group for each parameter. The proportion of patients with PCSA values at any post-baseline visit will also be displayed by duration of exposure for each treatment group.

Time to onset of the initial ALT and AST elevation (>3 x ULN) and total bilirubin elevation (>2 x ULN) (time to first observation of ALT > 3 x ULN or total bilirubin > 2 x ULN, whichever comes first) will be analyzed using Kaplan-Meier estimates, presented by treatment group. Consideration should be given to the impact of the spacing of scheduled tests. A graph of distribution of peak values of ALT versus peak values of total bilirubin will also be presented. Note that the ALT and total bilirubin values are presented on a logarithmic scale. The graph will be divided into 4 quadrants with a vertical line corresponding to 3 x ULN for ALT and a horizontal line corresponding to 2 x ULN for total bilirubin.

Listing of possible Hy's law cases identified by treatment group (eg, patients with any elevated ALT>3 x ULN, and associated with an increase in bilirubin ≥2 x ULN) with ALT, AST, alkaline phosphatase, total bilirubin, and the following complementary parameters: conjugated bilirubin and creatine phosphokinase, serum creatinine, complete blood count, HCV RNA.

Summarize the normalization by parameter (to ≤1 x ULN or return to baseline) of elevated liver function tests by categories of elevation (3 x, 5 x, 10 x, 20 x ULN for ALT and AST, 1.5 x ULN for alkaline phosphatase, and 1.5 x and 2 x ULN for total bilirubin), with the following categories of normalization: never normalized, normalized despite treatment continuation of IMP, or normalized after IMP discontinuation. Note that a patient will be counted only under the maximum elevation category.

2.4.5.4 Analyses of vital sign variables

The summary statistics (including number, mean, median, Q1, Q3, standard deviation, minimum and maximum) of all vital signs variables (values and changes from baseline) will be calculated for each visit or study assessment (baseline, each post-baseline time point, last on-treatment and/or worst on-treatment value) by treatment group. For all parameters, mean changes from baseline with the corresponding standard error will be plotted over time (at same time points) in each treatment group.

The incidence of PCSAs at any time during the treatment-emergent adverse event period will be summarized by treatment group irrespective of the baseline level and/or according to the following baseline status categories:

- Normal/missing
- Abnormal according to PCSA criterion or criteria

The incidence of PCSAs from baseline to Week 24 during the treatment-emergent adverse event period will be summarized by biological function and treatment group whatever the baseline level and/or according to the following baseline status categories:

- Normal/missing
- Abnormal according to PCSA criterion or criteria

2.4.5.5 Analyses of electrocardiogram variables

The incidence of PCSAs at any time during the treatment-emergent adverse event period will be summarized by treatment group irrespective of the baseline level and/or according to the following baseline status categories:

- Normal/missing
- Abnormal according to PCSA criterion or criteria

The incidence of PCSAs from baseline to Week 24 during the treatment-emergent adverse event period will be summarized by biological function and treatment group whatever the baseline level and/or according to the following baseline status categories:

- Normal/missing
- Abnormal according to PCSA criterion or criteria

2.4.6 Analyses of pharmacokinetic and pharmacodynamic variables

2.4.6.1 Pharmacokinetic analyses

2.4.6.1.1 Analyses of serum concentrations of SAR231893 (REGN668)

Serum concentrations of SAR231893 (REGN668) will be summarized in the PK population using arithmetic and geometric means, SD, standard error of the mean (SEM), coefficient of variation (CV), minimum, median and maximum per sampling time. If date and/or time of the drug injection and/or sampling is missing then the concentration will not be taken into account. For drug-treated patients, where concentration values are below the lower limit of quantification (LLOQ), one-half of the LLOQ will be used. Values will be expressed in the tables with no more than three significant figures. For patients in the placebo group, concentration values are below the LLOQ will be taken into account with a plasma concentration considered equal to 0.

2.4.6.1.2 Analyses of ADA variables

The following summary will be provided based on ADA population:

- Number (%) of patients with pre-existing immunoreactivity
- Number (%) of patients with treatment-emergent ADA
- The summary statistics (including number, median, Q1, Q3, minimum and maximum) of the peak post-baseline titer for patients with treatment-emergent ADA, and patients with persistent, indeterminate and transient ADA response
- Number (%) of patient with transient treatment-emergent ADA
- Number (%) of patients with persistent treatment-emergent ADA
- Number (%) of patients with indeterminate treatment-emergent ADA
- Number (%) of patients with treatment-boosted ADA
- The summary statistics (including number, median, Q1, Q3, minimum and maximum) of the peak post-baseline titer for patients with treatment-boosted ADA
- The summary statistics (including number, mean, SD, median, Q1, Q3, minimum and maximum) of the ratio of peak post-baseline titer to baseline titer for patients with treatment-boosted ADA
- Listing of ADA peak titer levels and neutralizing antibody status

- Number(%) of patients with neutralizing antibody status

Kinetics of treatment-emergent ADA response

Number (%) of patients with treatment-emergent ADA positive response at each visit will be summarized by each treatment group.

Plot of percentage of patients with treatment-emergent ADA positive response at each visit will be provided by each treatment group.

Impact of ADA on PK

Associations between ADA variables (eg, ADA peak titers, neutralizing antibody status, treatment-emergent, persistent, indeterminate and transient response, treatment-boosted) and serum concentration of dupilumab may be explored for dupilumab dose group. Plot of serum concentration of functional SAR231893 (REGN668) versus visit will be provided by ADA classifications for dupilumab dose group. Individual patient plots of PK according to ADA status will be provided to determine which individuals may have had PK impacted by ADAs.

Association of ADA with clinical efficacy endpoints

Associations between the ADA variables (eg, ADA peak titers, neutralizing antibody status, treatment-emergent, persistent, indeterminate and transient response, treatment-boosted) and the co-primary efficacy endpoints may be explored for the dupilumab dosed group.

Association of ADA with clinical safety endpoints

Association of safety versus ADA status may be analyzed in the ADA population. The safety assessment may focus on the following events:

- Severe injection site reactions last longer than 24 hours or serious injection site reactions.
- Hypersensitivity reactions (SMQ hypersensitivity narrow search confirmed by medical review)
- Anaphylactic reactions (SMQ anaphylactic reaction narrow search)
- In response to AESI like anaphylaxis or hypersensitivity additional ADA samples closer to the event may be analyzed, based on the judgment of the medical investigator and/or medical monitor.
- Associations between ADA variables (eg, ADA peak titers, neutralizing antibody status, treatment-emergent and treatment-boosted) and safety may be explored.

2.4.6.2 Pharmacodynamic [redacted] analyses

All biomarkers listed in Section 2.1.6 will be summarized in the Safety population defined as patients who actually received at least 1 dose or part of a dose of the IMP. Baseline values will be the last value collected prior to the first IMP. For randomized but not treated patients, baseline

value will be the last value collected prior to the randomization. Descriptive statistics (including number, mean, SD, median, Q1, Q3, min, max) of biomarkers at baseline will be summarized.

Summary plots (mean +/- standard error of the mean) on values at each visit, absolute changes from baseline and percent changes from baseline will be provided for each biomarker by treatment group.

[redacted]

2.4.7 Analyses of health economics variables

Analyses of healthcare resource utilization will be performed under the responsibility of the Health Economics Outcomes Research (HEOR) department of Sanofi. Methods and results will be made available in a separate report.

2.5 DATA HANDLING CONVENTIONS

2.5.1 General conventions

The following formulas will be used for computation of parameters.

Demographic formulas
Age is calculated as:

Integer part of (informed consent date -birth date)/365.25

Age of onset of nasal polyposis is calculated as:

Integer part of (nasal polyposis diagnosis date - birth date)/365.25

Age of onset of asthma is calculated as:

Integer part of (first diagnosis of asthma date - birth date)/365.25

BMI is calculated as:

Weight in kg / (height^2 in meters)

Renal function formulas

For patients ≥18 years old, creatinine clearance (CLcr) value will be derived using the equation of Cockroft and Gault:

$$\text{CLcr (ml/min)} = (140 - \text{age}) \times \text{weight (kg)} \times (1 - 0.15 \times \text{sex (0-M, 1-F)}) / (0.814 \times \text{creatinine } (\mu\text{mol/l}))$$

CLcr will be calculated using the last weight measurement on or before the visit of the creatinine measurement and age at the lab sampling date. Here age is calculated as following:

Age = integer part of (lab sampling date - birth date)/365.25

2.5.2 Data handling conventions for secondary efficacy variables

Periodical average of daily efficacy endpoints at designated study days

For the daily efficacy endpoints, the time period used to calculate the periodical average at each designated study day is summarized in Table 2. Randomization day is used as the reference day (Day 1).

Table 2 - Periodical average of daily efficacy assessment

Time point	Morning TSS and its components (nasal congestion, loss of smell, anterior rhinorrhea, posterior rhinorrhea)	Morning NPIF
Day 29	2-29	2-29
Day 57	30-57	30-57
Day 113	86-113	86-113
Day 169	142-169	142-169
Day 253	226-253	-
Day 337	310-337	-

2.5.3 Missing data

For categorical variables, patients with missing data are not included in calculations of percentages unless otherwise specified. When relevant, the number of patients with missing data is presented.

Handling of computation of treatment duration if investigational medicinal product end of treatment date is missing

For the calculation of the treatment duration, the date of the last dose of IMP is equal to the date of last administration reported on the end-of-treatment case report form page. If this date is missing, the exposure duration should be left as missing.

The last dose intake should be clearly identified in the case report form and should not be approximated by the last returned package date.

Handling of medication missing/partial dates

No imputation of medication start/end dates or times will be performed. If a medication date or time is missing or partially missing and it cannot be determined whether it was taken prior or concomitantly, it will be considered a prior, concomitant, and post-treatment medication.

Handling of adverse events with missing or partial date/time of onset

Missing or partial adverse event onset dates and times will be imputed so that if the partial adverse event onset date/time information does not indicate that the adverse event started prior to treatment or after the treatment-emergent adverse event period, the adverse event will be classified as treatment-emergent. No imputation of adverse event end dates/times will be performed. These data imputations are for categorization purpose only and will not be used in listings. No imputation is planned for date/time of adverse event resolution.

Handling of adverse events when date and time of first investigational medicinal product administration is missing

When the date and time of the first IMP administration is missing, all adverse events that occurred on or after the day of randomization should be considered as treatment-emergent adverse events. The exposure duration should be kept as missing.

The last dose intake should be clearly identified in the case report form and should not be approximated by the last returned package date.

Handling of missing assessment of relationship of adverse events to investigational medicinal product

If the assessment of the relationship to IMP is missing, then the relationship to IMP has to be assumed and the adverse event considered as such in the frequency tables of possibly related adverse events, but no imputation should be done at the data level.

Handling of missing severity of adverse events

If the severity is missing for 1 of the treatment-emergent occurrences of an adverse event, the maximal severity on the remaining occurrences will be considered. If the severity is missing for all the occurrences, a “missing” category will be added in the summary table.

Handling of potentially clinically significant abnormalities

If a patient has a missing baseline he will be grouped in the category “normal/missing at baseline.”

For PCSAs with 2 conditions, one based on a change from baseline value or a normal range and the other on a threshold value, with the first condition being missing, the PCSA will be based only on the second condition.

For a PCSA defined on a threshold and/or a normal range, this PCSA will be derived using this threshold if the normal range is missing; eg, for eosinophils the PCSA is > 0.5 GIGA/L or >ULN if ULN ≥0.5 GIGA/L. When ULN is missing, the value 0.5 should be used.

Measurements flagged as invalid by the laboratory will not be summarized or taken into account in the computation of PCSA values.

2.5.4 Windows for time points

For the safety assessment, the reference date for the derivation of relative days of events or findings will be the date of first IMP administration. Selected safety variables will be summarized by the analysis window define in Table 3 for the by visit descriptive analysis. All available values from central lab will be assigned to the appropriate visit window. In the event of multiple measurements of the same test in the same window, the one closest to the targeted visit date will be used for the by-visit summary. If they are at the same distance to the target day, the latest one will be used. For procedures planned on Visit 2, if it is done on the same date as the first IMP injection but the performance time is missing, it will belong to Visit 2 time window.

Table 3 – Time window for safety endpoints

		Time windows for					
Visit	**Target Day**	**Vital signs**	**Clinical lab testing**	**Serum Pregnancy test ,Hepatitis screen, HIV screen, ANA and Anti-ds DNA antibody**	**Urine Pregnancy test**	**Hepatitis B viral load**	**ECG**
Visit 1 (Week -4±1)	-28±7	<-14	<-14	1-			
Visit 2 (Week 0)	1	-14-1-	-14-1-		1-	1-	1-
Visit 3 (Week 2)	15						
Visit 4 (Week 4)	29				1+-42		
Visit 5 (Week 6)	43						
Visit 6 (Week 8)	57	1+-84			43-70		
(Week 12)	85				71-98		

		Time windows for					
Visit	Target Day	Vital signs	Clinical lab testing	Serum Pregnancy test ,Hepatitis screen, HIV screen, ANA and Anti-ds DNA antibody	Urine Pregnancy test	Hepatitis B viral load	ECG
Visit 7 (Week 16)	113	85-140	1^{+}-140		99-126		
(Week 20)	141				127-154		
Visit 8 (Week 24)	169	141-210	>140		155-182	1^{+}	1^{+}
(Week 28)	197				183-210		
(Week 32)	225				211-238		
Visit 9 (Week 36)	253	211-294			239-266		
(Week 40)	281				267-294		
(Week 44)	309				295-322		
Visit 10 (Week 48)	337	>294			>322		

1^{-}: up to 1st dose date/time; 1^{+}: after 1st dose date/time;

For the efficacy assessment, the reference date for the derivation of relative days of events or findings will be the randomization day. If a patient receives IMP prior to the randomization by mistake, the reference date of efficacy assessment will be the date of the first IMP administration for that patient. For the primary analyses, all available values of scheduled measurements will be assigned to the appropriate visit window according to Table 4. In the event of multiple measurements of the same test in the same window, the one closest to the targeted visit date will be used for the by-visit summary. If they are at the same distance to the target day, the latest one will be used. For the on-treatment sensitivity analyses, only scheduled measurements collected during the treatment epoch will be assigned to a time window.

Table 4 – Time window for efficacy variables

Visit	Target Day	Time windows for								
		Nasal endo-scopy	CT scan	UPSIT	VAS for rhino-sinusitis	sense of taste	Spirometry, ACQ-6	EQ-5D	SNOT-22	SF-36
Visit 1 (Week -4±1)	-28±7	<-14								
Visit 2 (Week 0)	1	-14-1^{-}	1^{-}	1^{-}	1^{-}	1^{-}	1^{-}	1^{-}	1^{-}	1^{-}
Visit 3 (Week 2)	15			1+-35	1+-21	1+-21				
Visit 4 (Week 4)	29			36-84	22-42	22-42				
Visit 5 (Week 6)	43									
Visit 6 (Week 8)	57	1+-84			43-84	43-84	1+-84		1+-84	
Visit 7 (Week 16)	113	85-140		85-140	85-140	85-140	85-140		85-140	
Visit 8 (Week 24)	169	141-210	1^{+}-252	141-252	141-210	>141	141-252	1^{+}	141-210	
Visit 9 (Week 36)	253	211-294			211-294				211-294	
Visit 10 (Week 48)	337	>294	>252	>252	>294		>252		>294	

1^{-}: up to randomization and before 1st dose date/time; 1^{+}: after randomization or 1st dose date/time

For the pharmacokinetics/pharmacodynamics variables summary, the reference date for the derivation of relative days of measurements will be the date of first IMP administration if the patient is treated with study treatment, or the randomization date if the patient is not treated. Pharmacokinetics /pharmacodynamics variables will be summarized by the analysis window defined in Table 5 for the by visit descriptive analyses. All available values of measurements will be assigned to the appropriate visit window. In the event of multiple measurements of the same test in the same window, the one closest to the targeted visit date will be used for the by-visit summary. If they are at the same distance to the target day, the latest one will be used. For procedures planned on Visit 2, if it is done on the same date as the first IMP injection but the performance time is missing, it will belong to Visit 2 time window.

Table 5 – Time window for pharmacokinetics/pharmacodynamics variables

Visit	Target day	Time windows for			
		TARC, Eotaxin-3, periostin,	total IgE, antigen-specific IgE	Serum dupilumab concentration	Anti-drug antibodies
Visit 1 (Week -4±1)	-28±7				
Visit 2 (Week 0)	1	1^-	1^-	1^-	1^-
Visit 3 (Week 2)	15				
Visit 4 (Week 4)	29			1^+-42	
Visit 5 (Week 6)	43				
Visit 6 (Week 8)	57			43-84	1^+-84
Visit 7 (Week 16)	113			85-140	85-140
Visit 8 (Week 24)	169		1^+	141-210	141-210
Visit 9 (Week 36)	253			211-294	211-294
Visit 10 (Week 48)	337			>294	>294

1^-: up to 1st dose date/time or randomization if patient is not treated; 1^+: after 1st dose date/time or randomization date if patient is not treated;

2.5.5 Unscheduled visits

Unscheduled visit measurements of laboratory data, vital signs, and ECG will be included in the by-visit summaries, computation of baseline, worst values, and PCSAs.

2.5.6 Pooling of centers for statistical analyses

Due to the large number of centers, the randomization is stratified by country. Due to small sample size in some countries, the countries will be pooled into regions as defined below for the analyses:

- East Europe: Russia, Ukraine, Cz. Republic, Hungary, Poland, Romania, Bulgaria
- Western Countries: France, Germany, Italy, United Kingdom, Netherlands, USA.

2.5.7 Statistical technical issues

NA

3 INTERIM ANALYSIS

No interim analysis will be performed before the final database lock described in Section 4.

4 DATABASE LOCK

The database lock is planned based on when all patients complete Week 48 visit (last patient last visit) or discontinue from the study before Week 48. Analyses will be based on all data collected up to this database lock and will be considered as the final analyses in the CSR.

5 SOFTWARE DOCUMENTATION

All summaries and statistical analyses will be generated using SAS version 9.4 or higher.

6 REFERENCES

1. Fokkens W, Lund VJ, Mullol J, Bachert C, Alobid I, Baroody F, Cohen N, et al. EPOS 2012: European position paper on rhinosinusitis and nasal polyps 2012. Rhinology. 2012;50:1-12.

2. Gevaert P, Calus L, Van Zele T, Blomme K, De Ruyck N, Bauters W, et al. Omalizumab is effective in allergic and nonallergic patients with nasal polyps and asthma. J Allergy Clin.Immunol;2013;131:110-6

3. Gevaert P, Van Bruaene N, Cattaert T, Van Steen K., Van Zele T, Acke F, et al. Mepolizumab, a humanized anti-IL-5 mAb, as a treatment option for severe nasal polyposis. J Allergy Clin Immunol. 2011;128:989-95.

4. Van Zele T, Gevaert P, Holtappels G, Beule A, Wormald PJ, Mayr S, et al. Oral steroids and doxycycline: two different approaches to treat nasal polyps. J Allergy Clin Immunol. 2010;125:1069-76.

5. Lund VJ, Mackay, IS. Staging in rhinosinusitus. Rhinology. 1993;31:183-4.

6. Metson R, Gliklich RE, Stankiewicz JA, Kennedy DW, Duncavage JA, Hoffman SR, et al. Comparison of sinus computed tomography staging systems. Otolaryngol Head Neck Surg. 1997;117:372-9.

7. Oluwole M, Russell N, Tan L, Gardiner Q, White P. A comparison of computerized tomographic staging systems in chronic sinusitis. Clin Otolaryngol Allied Sci. 1996; 21: 91-95.

8. Hopkins C, Gillett S, Slack R, Lund VJ, Browne JP. Psychometric validity of the 22- item sinonasal outcome test. Clin Otolaryngol. 2009;34:447-54.

9. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. European Respiratory Journal. 2005; 26:319-338.

10. Juniper EF, Svensson K, Mork AC, Stahl E. Measurement properties and interpretation of three shortened versions of the asthma control questionnaire. Respir Med. 2005;99:553-8.

11. Van Agthoven M, Fokkens WJ, Van de Merwe JP, Marijke Van Bolhuis E, Uyl-de Groot CA, Busschbach JJ. Am J Rhinol. 2001;15:231-7.

12. Carpenter JR, Roger JH, Kenward MG. Analysis of Longitudinal Trials with Protocol Deviation: A Framework for Relevant, Accessible Assumptions, and inference via Multiple Imputation. Journal of Biopharmaceutical Statistics. 2013; 23:1352-1371.

13. Fan C. On analysis of binary response data in longitudinal factorial studies. Journal of Statistical Computation and Simulation 2017; 87(1): 100-122.